



基于对象更新的邻域多粒度粗糙集模型增量式算法

陈宝国, 邓明

引用本文:

陈宝国, 邓明. 基于对象更新的邻域多粒度粗糙集模型增量式算法[J]. 智能系统学报, 2023, 18(3): 562–576.

CHEN Baoguo, DENG Ming. An incremental algorithm for the neighborhood multi-granulation rough set model based on object update[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(3): 562–576.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202112042>

您可能感兴趣的其他文章

基于广义自适应多粒度的多源信息融合研究

Multi-source information fusion through generalized adaptive multi-granulation
智能系统学报. 2023, 18(1): 173–185 <https://dx.doi.org/10.11992/tis.202208030>

基于对象变化的邻域决策粗糙集动态更新算法

Dynamic updating algorithm of neighborhood decision-theoretic rough set model based on object change
智能系统学报. 2021, 16(4): 746–756 <https://dx.doi.org/10.11992/tis.202010028>

不完备数据中面向特征值更新的增量特征选择方法

Incremental approach for feature selection in incomplete data while updating feature values
智能系统学报. 2021, 16(3): 493–501 <https://dx.doi.org/10.11992/tis.202006045>

变精度下不完备邻域决策系统的属性约简算法

Attribute reduction algorithm of the incomplete neighborhood decision system with variable precision
智能系统学报. 2017, 12(3): 386–391 <https://dx.doi.org/10.11992/tis.201705027>

基于证据理论刻画多粒度覆盖粗糙集的数值属性

Evidence-theory-based numerical characterization of multi-granulation covering rough sets
智能系统学报. 2016, 11(4): 481–486 <https://dx.doi.org/10.11992/tis.201606011>

元素最小描述并集下的多粒度覆盖粗糙集模型

Multigranulation covering rough sets based on the union of minimal descriptions of elements
智能系统学报. 2016, 11(4): 534–538 <https://dx.doi.org/10.11992/tis.201605034>

DOI: 10.11992/tis.202112042

网络出版地址: <https://kns.cnki.net/kcms2/detail/23.1538.TP.20230320.1227.002.html>

基于对象更新的邻域多粒度粗糙集模型增量式算法

陈宝国, 邓明

(淮南师范学院 计算机学院, 安徽 淮南 232038)

摘要: 邻域多粒度粗糙集模型是粗糙集理论的重要研究分支。然而在大数据环境下, 数据时刻处于动态更新之中, 针对数值型信息系统对象动态变化的情形, 本文提出一种邻域多粒度粗糙集模型的增量式更新算法。文中首先利用矩阵的方法表示了邻域多粒度粗糙集中邻域类与目标近似集之间的两种近似关系, 分别称之为子集近似关系矩阵和交集近似关系矩阵, 并通过这两种近似关系矩阵重构了邻域多粒度粗糙集模型; 然后针对数值型信息系统对象增加和对象减少的情形, 研究了这两种近似关系矩阵随对象变化时的增量式更新, 理论分析证明了这种更新方法的高效性; 最后基于近似关系矩阵的增量式更新设计出了邻域多粒度粗糙集模型的增量式更新算法。实验结果验证了所提出增量式算法的有效性和优越性。

关键词: 数据更新; 粗糙集; 多粒度; 邻域; 对象变化; 增量式学习; 近似关系矩阵; 增量式算法

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2023)03-0562-15

中文引用格式: 陈宝国, 邓明. 基于对象更新的邻域多粒度粗糙集模型增量式算法 [J]. 智能系统学报, 2023, 18(3): 562-576.

英文引用格式: CHEN Baoguo, DENG Ming. An incremental algorithm for the neighborhood multi-granulation rough set model based on object update[J]. CAAI transactions on intelligent systems, 2023, 18(3): 562-576.

An incremental algorithm for the neighborhood multi-granulation rough set model based on object update

CHEN Baoguo, DENG Ming

(School of Computer Science, Huainan Normal University, Huainan 232038, China)

Abstract: The neighborhood multi-granulation rough set model is an important branch of rough set theory. However, in the big data environment, the data is constantly being updated dynamically. In view of the dynamic changes of numerical information system objects, an incremental updating algorithm for neighborhood multi-granulation rough set model is proposed in this paper. Firstly, two kinds of approximation relations between neighborhood class and target approximation set in the neighborhood multi-granulation rough sets are expressed by the matrix method, which are called subset approximation relation matrix and intersection approximation relation matrix, respectively. The neighborhood multi-granulation rough set model is reconstructed by these two approximation relation matrices. Then, the incremental updating of these two approximation relation matrices is studied in the case of increasing and decreasing objects in numerical information system. The theoretical analysis proves that this updating method is of high-efficiency. Finally, the incremental updating algorithm of neighborhood multi-granulation rough set model is designed based on the incremental updating of approximation relation matrices. Experimental results verify the effectiveness and superiority of the proposed incremental algorithm.

Keywords: data update; rough set; multi-granulation; neighborhood; object change; incremental learning; approximation relation matrix; incremental algorithm

粗糙集理论是波兰学者 Pawlak 提出的一种

数据挖掘模型^[1], 由于它在不精确和不确定性的数据环境下发挥出了重要的作用, 目前已被大规模用于机器学习和数据挖掘等^[2-3]领域中。

传统的粗糙集模型通过单个的二元关系对数

收稿日期: 2021-12-30. 网络出版日期: 2023-03-21.

基金项目: 安徽省高校自然科学研究重点项目 (KJ2018A0469).

通信作者: 陈宝国. E-mail: bgchen0706@163.com.

据进行分割粒化,从而完成对近似目标的粗糙逼近。近年来,为了从多个粒度层次对目标信息进行全方位的挖掘,Qian等^[4]将传统的单个二元关系延伸到多个二元关系,提出了多粒度粗糙集模型,该模型的提出,使得粗糙集理论在分布式数据和多粒化数据的处理上有了一种新的解决方案。在此基础上,学者们对Qian的多粒度粗糙集模型进行了不断地推广和改进,例如将多粒度粗糙集与决策粗糙集结合,提出了多粒度决策粗糙集模型^[5]。将多粒度粗糙集与三支决策模型结合,提出了多粒度三支决策模型^[6],针对属性和属性值的粒化,Xu^[7]提出了一种新型的改进多粒度粗糙集模型。为了进一步扩大多粒度粗糙集的适用范围,Lin等^[8]将邻域关系融入了多粒度粗糙集模型中,提出了邻域多粒度粗糙集,使得该模型成为了多粒度粗糙集研究分支的热点内容^[9-11]。

当前大数据环境下,数据时刻处于动态变化之中,传统的机器学习和数据挖掘算法针对这一情形面临着极大的挑战。在粗糙集理论中,通过对模型和算法构造增量式学习可以大幅度改善在动态数据下的处理效率,提高粗糙集理论的实际应用性能。增量式学习即保留旧数据的模型和算法结果,然后在此基础上对变化的数据进行计算,使得达到最终新数据的增量式更新。例如,Luo等^[12]利用矩阵方法构造了决策粗糙集模型的增量式更新;Wei等^[13]基于区分矩阵的增量式更新构造了一种属性约简算法;Shu等^[14]研究了邻域粗糙集的增量式学习,并提出一种增量式属性约简;Huang等^[15]利用矩阵的方法重构了多源数据的条件熵模型,并用于增量式属性约简的设计;沈玉峰^[16]基于区分度的矩阵表达提出了一种增量式属性约简算法;Yang等^[17]总结性的提出了一种增量式属性约简的统一框架;Huang等^[18]研究了邻域三支决策模型的增量式更新,并设计出了一种增量式更新算法;杨臻等^[19]利用集合近似的方法针对混合型变精度粗糙集提出了一种动态更新算法;刘丹等^[20]提出了一种双论域模糊概率粗糙集的增量式更新算法。针对多粒度粗糙集模型,学者们也进行了相关的增量式学习的研究,例如Yang等^[21]基于集合近似的方法研究了多粒度粒结构变化时,多粒度粗糙集模型的增量式更新,在其基础上,Hu等^[22]进行进一步地优化和改进,利用矩阵方法去表达多粒度粗糙集,并同样设计出了一种粒结构变化情形下的增量式更新算法,根据这一方法,Hu等^[23]进一步设计出了一种优势多粒度粗糙集的增量式更新算法。同时对于邻域多粒度粗糙集模型,Hu等^[24]也针对粒结构

动态变化的情形提出了一种增量式更新算法,并且徐怡等^[25]也提出了类似的改进算法。

然而,针对多粒度粗糙集模型计算的增量式学习研究,目前大多都是考虑粒结构变化的情形,而对象的变化也是信息系统一种较为常见的更新形式,对于多粒度粗糙集而言这一方面的研究目前还比较少。在文献[26]中,两位学者利用矩阵的方法提出了对象变化情形下邻域多粒度粗糙集模型的增量式更新,其主要思想是利用邻域关系矩阵去计算模型的上下近似集,当论域中对象发生变化时,直接增量式更新每个粒度的邻域关系矩阵,同时重新利用矩阵结构去计算新的上下近似结果。这使得文献[26]所提出的增量式方法具有大量的矩阵运算,并且存在着一定的不必要计算,因此仍有一定的优化空间。

为了进一步提高邻域多粒度粗糙集模型的增量式更新效率,本文针对信息系统对象变化的情形,提出一种改进的邻域多粒度粗糙集模型增量式更新方法。矩阵是构造增量式学习的一种重要工具^[12-13,15-16,20,22,24-26],与文献[26]类似,本文也同样采用矩阵的方法去构造模型的增量式学习。文中首先在邻域多粒度粗糙集模型中提出两种近似关系矩阵的概念,然后针对信息系统对象增加和对象减少的情形,在旧信息系统两种近似关系矩阵的基础上进一步地增量式更新出新信息系统的两种近似关系矩阵,最后根据两种近似关系矩阵的增量式更新方法,提出了一种邻域多粒度粗糙集模型的增量式更新算法。本文所提出的增量式更新方法与文献[26]的不同之处在于,本文提出的两种近似关系矩阵表示的是对象在各个粒度下邻域类与近似目标集之间的集合隶属近似关系,文中理论分析表明,通过这两种近似关系矩阵可以直接快速地得到邻域多粒度粗糙集的上下近似集结果,其进行更新的计算方式以及计算的效率均优于文献[26],最终实验结果也证明了所提出算法的有效性和优越性。

1 基本理论

本节主要介绍多粒度粗糙集和邻域多粒度粗糙集的相关基本理论,为后文的研究提供基础。

在粗糙集理论中,数据集被表示成信息系统 $S = (U, A_T)$ 的形式,其中 U 称为信息系统的论域,是由数据样本组成的集合, A_T 称为信息系统的全体属性集,即数据集所有特征的集合。对于论域中的对象 $\forall x \in U$ 在属性 $\forall a \in A_T$ 下的属性值表示为 $a(x)$,若属性值 $a(x)$ 均为离散值类型,那么该信息系统称之为离散型信息系统,若属性值 $a(x)$ 均为数值

类型,那么该信息系统称之为数值型信息系统。

定义 1^[1] 给定离散型信息系统 $S = (U, A_T)$, 属性子集 $A \subseteq A_T$, 那么由属性子集 A 确定的等价关系定义为 $E_A = \{(x, y) \in U \times U | \forall a \in A, a(x) = a(y)\}$ 。对于离散型信息系统论域中的对象 $\forall x \in U$, 在等价关系 E_A 下确定出的等价类定义为 $[x]_A = \{y \in U | (x, y) \in E_A\}$ 。

传统的等价关系仅适用于离散型的信息系统,然而实际应用中存在着数值型的信息系统,因此学者们将等价关系进行推广,提出了邻域关系以及邻域粗糙集模型。

定义 2^[8] 给定数值型信息系统 $S = (U, A_T)$, 属性子集 $A \subseteq A_T$, 设邻域半径为 δ , 那么由属性子集 A 确定的邻域关系定义为

$$N_A^\delta = \{(x, y) \in U \times U | d_A(x, y) \leq \delta\}$$

式中: 邻域半径 δ 是一个非负常数, $d_A(x, y)$ 表示对象 x 和 y 在属性子集 A 下的距离度量, 通常 $d_A(x, y)$ 定义为 $d_A(x, y) = \left(\sum_{a \in A} |a(x) - a(y)|^p\right)^{1/p}$ 。这里的 p 一般取 1、2 或 ∞ 。对于 $\forall x \in U$ 在邻域关系 N_A^δ 下的邻域类定义为 $\delta_A(x) = \{y \in U | (x, y) \in N_A^\delta\}$ 。下文中, 在不引起混淆的情况下, 将 N_A^δ 简单表示为 N_A 。

传统的粗糙集模型通过单一粒度视角对近似目标集进行粗糙近似, Qian 等^[4] 在其基础上推广至多粒度视角下, 提出了基于等价关系的多粒度粗糙集模型, 同时, Lin 等^[8] 又将等价关系推广至邻域关系, 提出了邻域多粒度粗糙集模型。

定义 3^[8] 给定数值型信息系统 $S = (U, A_T)$, 设 A_1, A_2, \dots, A_m 是属性集 A_T 的一个属性子集族, 其中 $A_i \subseteq A_T, 1 \leq i \leq m$, 邻域半径为 δ , 属性子集 A_i 确定的邻域关系为 N_{A_i} , 那么对象集 $X \subseteq U$ 在属性子集族 A_1, A_2, \dots, A_m 下的乐观邻域多粒度下近似集和乐观邻域多粒度上近似集分别定义为

$$\begin{aligned} \sum_{i=1}^m N_{A_i}^o(X) &= \{x \in U | \delta_{A_i}(x) \subseteq X \vee \\ &\quad \delta_{A_2}(x) \subseteq X \vee \dots \vee \delta_{A_m}(x) \subseteq X\} \\ \sum_{i=1}^m N_{A_i}^u(X) &= \{x \in U | \delta_{A_i}(x) \cap X \neq \emptyset \wedge \\ &\quad \delta_{A_2}(x) \cap X \neq \emptyset \wedge \dots \wedge \delta_{A_m}(x) \cap X \neq \emptyset\} \end{aligned}$$

同时对象集 $X \subseteq U$ 在属性子集族 A_1, A_2, \dots, A_m 下的悲观邻域多粒度下近似集和悲观邻域多粒度上近似集分别定义为

$$\begin{aligned} \sum_{i=1}^m N_{A_i}^p(X) &= \{x \in U | \delta_{A_i}(x) \subseteq X \wedge \\ &\quad \delta_{A_2}(x) \subseteq X \wedge \dots \wedge \delta_{A_m}(x) \subseteq X\} \\ \sum_{i=1}^m N_{A_i}^b(X) &= \{x \in U | \delta_{A_i}(x) \cap X \neq \emptyset \vee \\ &\quad \delta_{A_2}(x) \cap X \neq \emptyset \vee \dots \vee \delta_{A_m}(x) \cap X \neq \emptyset\} \end{aligned}$$

称 $\left\langle \sum_{i=1}^m N_{A_i}^o(X), \sum_{i=1}^m N_{A_i}^u(X) \right\rangle$ 为对象集 $X \subseteq U$ 在属性

子集族 A_1, A_2, \dots, A_m 下的乐观邻域多粒度粗糙近似

集, 称 $\left\langle \sum_{i=1}^m N_{A_i}^p(X), \sum_{i=1}^m N_{A_i}^b(X) \right\rangle$ 为对象集 $X \subseteq U$ 在属性子

集族 A_1, A_2, \dots, A_m 下的悲观邻域多粒度粗糙近似集。

2 对象更新时邻域多粒度粗糙集的增量式更新

由于现实中数据集的论域处于不断动态更新之中, 即对象的增加或对象的减少, 对于邻域多粒度粗糙集模型, 如果直接按照模型的定义计算上下近似集, 其计算效率不足以满足实际的需求。针对这一问题, 本节将设计出一种高效的邻域多粒度粗糙集增量式更新方法。

对于数值型信息系统 $S = (U, A_T)$ 以及属性集 A_T 构造的属性子集族 A_1, A_2, \dots, A_m , 在不引起混淆的情况, 用 $\varphi^o(X)$ ($\varphi^p(X)$) 表示近似目标对象集 $X \subseteq U$ 的乐观 (悲观) 邻域多粒度下近似集, 即 $\varphi^o(X) = \sum_{i=1}^m N_{A_i}^o(X)$ ($\varphi^p(X) = \sum_{i=1}^m N_{A_i}^p(X)$), 在不区分乐观和悲观的情形下, 用 $\varphi(X)$ 统一表示 $\varphi^o(X)$ 和 $\varphi^p(X)$ 。类似地, 用 $\phi^o(X)$ ($\phi^p(X)$) 表示近似目标对象集 $X \subseteq U$ 的乐观 (悲观) 邻域多粒度上近似集, 即 $\phi^o(X) = \sum_{i=1}^m N_{A_i}^u(X)$ ($\phi^p(X) = \sum_{i=1}^m N_{A_i}^b(X)$), 在不区分乐观和悲观的情形下, 用 $\phi(X)$ 统一表示用 $\phi^o(X)$ 和 $\phi^p(X)$ 。

2.1 基于近似关系矩阵形式的邻域多粒度粗糙集

Hu 等^[26] 通过矩阵的方法定义邻域关系, 然后利用邻域关系矩阵与近似目标集的特征向量进行矩阵运算, 重构了邻域多粒度粗糙集模型, 最后根据邻域关系矩阵的增量式更新进而更新邻域多粒度粗糙集模型, 然而 Hu 等的增量式方法其计算复杂度仍然较高, 其性能方面仍有一定的优化空间。本节这里不同于 Hu 等的方法, 将提出邻域多粒度粗糙集一种新的矩阵表达形式, 其中定义了邻域多粒度粗糙集的两种近似关系矩阵, 本文分别称之为子集近似关系矩阵和交集近似关系矩阵, 并利用这两种近似关系矩阵去重构邻域多粒度粗糙集模型。

定义 4 给定数值型信息系统 $S = (U, A_T)$, $U = \{x_1, x_2, \dots, x_n\}$, 设 A_1, A_2, \dots, A_m 是属性集 A_T 的一个属性子集族, 其中 $A_j \subseteq A_T, 1 \leq j \leq m$, 邻域半径为 δ , 属性子集 A_j 确定的邻域关系为 N_{A_j} , 对象 $\forall x_i \in U$ 在邻域关系 N_{A_j} 下的邻域类为 $\delta_{A_j}(x_i)$ 。给定近似目标对

象集 $X \subseteq U$, 定义该信息系统中 X 在 A_1, A_2, \dots, A_m 下的子集近似关系矩阵 $\Theta(X)$ 和交集近似关系矩阵 $\Omega(X)$ 分别为

$$\Theta(X) = \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1m} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{n1} & \theta_{n2} & \cdots & \theta_{nm} \end{bmatrix} = (\theta_{ij})_{n \times m}$$

$$\text{其中 } \theta_{ij} = \begin{cases} 1, & \delta_{A_j}(x_i) \subseteq X \\ 0, & \delta_{A_j}(x_i) \not\subseteq X \end{cases} \quad 1 \leq i \leq n, 1 \leq j \leq m.$$

$$\Omega(X) = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1m} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n1} & \omega_{n2} & \cdots & \omega_{nm} \end{bmatrix} = (\omega_{ij})_{n \times m}$$

$$\text{其中 } \omega_{ij} = \begin{cases} 1, & \delta_{A_j}(x_i) \cap X \neq \emptyset \\ 0, & \delta_{A_j}(x_i) \cap X = \emptyset \end{cases} \quad 1 \leq i \leq n, 1 \leq j \leq m.$$

定义4表明, 当数值型信息系统 $S = (U, A_T)$ 包含了 n 个对象, 并且邻域多粒度粗糙集模型构建了 m 个属性子集 A_1, A_2, \dots, A_m 的属性子集族, 那么邻域多粒度近似集 $X \subseteq U$ 的子集和交集近似关系矩阵具有 n 行 m 列, 其中矩阵第 i 行反映了对象 $x_i \in U$ 与近似目标集 X 在不同粒度下的集合隶属关系。假如当 $\delta_{A_p}(x_i) \subseteq X$ 、 $\delta_{A_q}(x_i) \subseteq X$ 且 $\delta_{A_r}(x_i) \subseteq X$, 那么对应的子集近似关系矩阵 $\Theta(X)$ 中第 i 行第 p 、 q 和 r 列上的元素为 1, 该行其他列的元素为 0。同样地, 假如当 $\delta_{A_s}(x_i) \cap X \neq \emptyset$ 和 $\delta_{A_t}(x_i) \cap X \neq \emptyset$, 那么对应的交集近似关系矩阵 $\Omega(X)$ 中第 i 行第 s 和 t 列上的元素为 1, 该行其他列的元素为 0。因此可以看出, 定义4中所定义的近似关系矩阵是对整个邻域多粒度粗糙集中邻域类与近似集之间细节关系的直观展示。下文中, 在不引起混淆的情况下, 将子集近似关系矩阵和交集近似关系矩阵统称为近似关系矩阵。

近似关系矩阵满足如下性质。

性质1 给定数值型信息系统 $S = (U, A_T)$ 和属性子集族 A_1, A_2, \dots, A_m , 近似目标对象集 $X \subseteq U$ 在 A_1, A_2, \dots, A_m 下的乐观邻域多粒度下近似集和上近似集分别为 $\varphi^o(X)$ 和 $\phi^o(X)$, 那么有:

1) 对于 $x_i, x_j \in U$, 若 $x_i \in \varphi^o(X)$, $x_j \notin \varphi^o(X)$, 那么子集近似关系矩阵 $\Theta(X)$ 中第 i 行至少存在一个元素为 1, 第 j 行所有元素均为 0。

2) 对于 $x_i, x_j \in U$, 若 $x_i \in \phi^o(X)$, $x_j \notin \phi^o(X)$, 那么交集近似关系矩阵 $\Omega(X)$ 中第 i 行所有元素均为 1, 第 j 行至少存在一个元素为 0。

证明 根据定义3关于乐观邻域多粒度粗糙集的定义以及定义4关于近似关系矩阵的定义:

1) 对于乐观邻域多粒度粗糙集下近似集 $\varphi^o(X)$ 和属性子集族 A_1, A_2, \dots, A_m , 若 $x_i \in \varphi^o(X)$, 那么 $\delta_{A_1}(x_i) \subseteq X \vee \delta_{A_2}(x_i) \subseteq X \vee \dots \vee \delta_{A_m}(x_i) \subseteq X$, 即至少存在一

个属性子集 A_t 满足 $\delta_{A_t}(x_i) \subseteq X$, 所以子集近似关系矩阵 $\Theta(X)$ 中第 i 行至少存在一个元素为 1。对于 $x_j \notin \varphi^o(X)$, 那么就满足关系 $\delta_{A_1}(x_j) \not\subseteq X \wedge \delta_{A_2}(x_j) \not\subseteq X \wedge \dots \wedge \delta_{A_m}(x_j) \not\subseteq X$, 即对象 x_j 均有 $\delta_{A_t}(x_j) \not\subseteq X (1 \leq t \leq m)$, 所以子集近似关系矩阵 $\Theta(X)$ 中第 j 行所有元素均为 0。

2) 对于乐观邻域多粒度粗糙集上近似集 $\phi^o(X)$ 和属性子集族 A_1, A_2, \dots, A_m , 若 $x_i \in \phi^o(X)$, 则 $\delta_{A_1}(x_i) \cap X \neq \emptyset \wedge \dots \wedge \delta_{A_m}(x_i) \cap X \neq \emptyset$, 即均有 $\delta_{A_t}(x_i) \cap X \neq \emptyset (1 \leq t \leq m)$, 所以交集近似关系矩阵 $\Omega(X)$ 中第 i 行所有元素均为 1。若 $x_j \notin \phi^o(X)$, 那么满足关系 $\delta_{A_1}(x_j) \cap X = \emptyset \vee \dots \vee \delta_{A_m}(x_j) \cap X = \emptyset$, 即存在属性子集 A_t 满足 $\delta_{A_t}(x_j) \cap X = \emptyset (1 \leq t \leq m)$, 所以交集近似关系矩阵 $\Omega(X)$ 中第 j 行至少存在一个元素为 0。

证毕。

性质2 给定数值型信息系统 $S = (U, A_T)$ 和属性子集族 A_1, A_2, \dots, A_m , 近似目标对象集 $X \subseteq U$ 在 A_1, A_2, \dots, A_m 下的悲观邻域多粒度下近似集和上近似集分别为 $\varphi^p(X)$ 和 $\phi^p(X)$, 那么有:

1) 对于 $x_i, x_j \in U$, 若 $x_i \in \varphi^p(X)$, $x_j \notin \varphi^p(X)$, 那么子集近似关系矩阵 $\Theta(X)$ 中第 i 行所有元素均为 1, 第 j 行至少存在一个元素为 0。

2) 对于 $x_i, x_j \in U$, 若 $x_i \in \phi^p(X)$, $x_j \notin \phi^p(X)$, 那么交集近似关系矩阵 $\Omega(X)$ 中第 i 行至少存在一个元素为 1, 第 j 行所有元素均为 0。

证明 证明过程同性质1。

性质3 给定数值型信息系统 $S = (U, A_T)$, 其中 $U = \{x_1, x_2, \dots, x_n\}$, 给定属性子集族 A_1, A_2, \dots, A_m , 近似目标对象集 $X \subseteq U$ 在 A_1, A_2, \dots, A_m 下的子集近似关系矩阵和交集近似关系矩阵分别为 $\Theta(X)$ 和 $\Omega(X)$, 那么 X 在 A_1, A_2, \dots, A_m 下的乐观邻域多粒度下近似集 $\varphi^o(X)$ 和上近似集 $\phi^o(X)$ 可分别表示为

$$\varphi^o(X) = \{x_i \in U (1 \leq i \leq n) |$$

$$\exists \theta_{ij} = 1 (1 \leq j \leq m), \Theta(X) = (\theta_{ij})_{n \times m}\}$$

$$\phi^o(X) = \{x_i \in U (1 \leq i \leq n) |$$

$$\forall \omega_{ij} = 1 (1 \leq j \leq m), \Omega(X) = (\omega_{ij})_{n \times m}\}$$

证明 由性质1可以直接得到证明。

性质4 给定数值型信息系统 $S = (U, A_T)$, 其中 $U = \{x_1, x_2, \dots, x_n\}$, 给定属性子集族 A_1, A_2, \dots, A_m , 近似目标对象集 $X \subseteq U$ 在 A_1, A_2, \dots, A_m 下的子集近似关系矩阵和交集近似关系矩阵分别为 $\Theta(X)$ 和 $\Omega(X)$, 那么 X 在 A_1, A_2, \dots, A_m 下的悲观邻域多粒度下近似集 $\varphi^p(X)$ 和上近似集 $\phi^p(X)$ 可分别表示为

$$\varphi^p(X) = \{x_i \in U (1 \leq i \leq n) |$$

$$\forall \theta_{ij} = 1 (1 \leq j \leq m), \Theta(X) = (\theta_{ij})_{n \times m}\}$$

$$\phi^p(X) = \{x_i \in U (1 \leq i \leq n) |$$

$$\exists \omega_{ij} = 1 (1 \leq j \leq m), \Omega(X) = (\omega_{ij})_{n \times m}\}$$

证明 由性质2可以直接得到证明。

性质 3 和性质 4 通过两种近似关系矩阵的形式对邻域多粒度粗糙集模型进行了重新表达,因此给定近似目标对象集以及近似目标对象集在多粒度下的近似关系矩阵,那么根据性质 3 和性质 4 可以直接得到邻域多粒度粗糙集模型的结果。

性质 3 和性质 4 所示的邻域多粒度粗糙集表示方法不同于文献 [26] 提出的方法,文献 [26] 为每个邻域关系都建立了邻域关系矩阵,其关系矩阵的建立以单个对象为视角,而定义 4 中近似关系矩阵的建立以对象的邻域类为视角,对于论域 $U = \{x_1, x_2, \dots, x_n\}$ 和属性子集族 A_1, A_2, \dots, A_m , 文献 [26] 需要构建 m 个 $n \times n$ 的矩阵,并且联合这 m 个矩阵计算模型的上下近似集结果,而本文仅需构建两个 $n \times m$ 的矩阵,每个矩阵对应计算出模型上下近似集结果,因此本文方法的计算规模和复杂程度都大幅度小于文献 [26],从而提升了其计算效率。

2.2 对象增加时邻域多粒度粗糙集的增量式更新

本小节中,在 2.1 节所提出的两种近似关系矩阵的基础上,进一步探究对象增加时邻域多粒度粗糙集的增量式更新方法。

通常信息系统对象的增加和减少都是批量进行的,而直接针对批量变化的对象进行增量式更新研究,其过程比较复杂,本文这里采用其他学者类似的研究方式,每次只研究单个对象变化时模型的增量式更新,然后针对多个对象通过逐步迭代的方式可以实现最终的增量式更新^[19]。

性质 3 和性质 4 已经表明,通过近似目标对象集在多粒度环境下的子集近似关系矩阵和交集近似关系矩阵,可以直接得到邻域多粒度粗糙集模型的结果,因此我们可以首先对子集近似关系矩阵和交集近似关系矩阵的增量式更新进行研究,然后依托于这两种近似关系矩阵的增量式更新,那么最终就可以直接实现邻域多粒度粗糙集模型结果的增量式更新。

定理 1 给定数值型信息系统 $S = (U, A_T)$, $U = \{x_1, x_2, \dots, x_n\}$, 设属性子集族 A_1, A_2, \dots, A_m , 邻域半径为 δ 。近似目标对象集 $X \subseteq U$ 在 A_1, A_2, \dots, A_m 下的子集近似关系矩阵和交集近似关系矩阵分别为 $\Theta(X) = (\theta_{ij})_{n \times m}$ 和 $\Omega(X) = (\omega_{ij})_{n \times m}$ 。当数值型信息系统论域增加对象 x_{n+1} , $U^+ = U \cup \{x_{n+1}\}$, 对象 x_{n+1} 在属性子集 $A_i (1 \leq i \leq m)$ 下关于论域 U^+ 的邻域类为 $\delta_{A_i}(x_{n+1})$ 。对于两种新的近似目标对象集 $X_1 = X$ 和 $X_2 = X \cup \{x_{n+1}\}$, 论域 U^+ 下 X_1 在 A_1, A_2, \dots, A_m 下的子集近似关系矩阵和交集近似关系矩阵分别为 $\Theta^+(X_1)$ 和 $\Omega^+(X_1)$, 论域 U^+ 下 X_2 在 A_1, A_2, \dots, A_m 下的子集近似关

系矩阵和交集近似关系矩阵分别为 $\Theta^+(X_2)$ 和 $\Omega^+(X_2)$, 那么它们可增量式计算为

$$\Theta^+(X_1) = \begin{bmatrix} \theta_{11}^+ & \theta_{12}^+ & \cdots & \theta_{1m}^+ \\ \theta_{21}^+ & \theta_{22}^+ & \cdots & \theta_{2m}^+ \\ \vdots & \vdots & \theta_{ij}^+ & \vdots \\ \theta_{n1}^+ & \theta_{n2}^+ & \cdots & \theta_{nm}^+ \\ \theta_{(n+1)1}^+ & \theta_{(n+1)2}^+ & \cdots & \theta_{(n+1)m}^+ \end{bmatrix} = (\theta_{ij}^+)_{(n+1) \times m}$$

其中

$$\begin{cases} \theta_{ij}^+ = 0, & x_i \in \delta_{A_j}(x_{n+1}), 1 \leq i \leq n, 1 \leq j \leq m \\ \theta_{ij}^+ = \theta_{ij}, & x_i \notin \delta_{A_j}(x_{n+1}), 1 \leq i \leq n, 1 \leq j \leq m \\ \theta_{(n+1)j}^+ = 0, & 1 \leq j \leq m \end{cases}$$

$$\Omega^+(X_1) = \begin{bmatrix} \omega_{11}^+ & \omega_{12}^+ & \cdots & \omega_{1m}^+ \\ \omega_{21}^+ & \omega_{22}^+ & \cdots & \omega_{2m}^+ \\ \vdots & \vdots & \omega_{ij}^+ & \vdots \\ \omega_{n1}^+ & \omega_{n2}^+ & \cdots & \omega_{nm}^+ \\ \omega_{(n+1)1}^+ & \omega_{(n+1)2}^+ & \cdots & \omega_{(n+1)m}^+ \end{bmatrix} = (\omega_{ij}^+)_{(n+1) \times m}$$

其中

$$\begin{cases} \omega_{ij}^+ = \omega_{ij}, & 1 \leq i \leq n, 1 \leq j \leq m \\ \omega_{(n+1)j}^+ = 1, & \delta_{A_j}(x_{n+1}) \cap X_1 \neq \emptyset, 1 \leq j \leq m \\ \omega_{(n+1)j}^+ = 0, & \delta_{A_j}(x_{n+1}) \cap X_1 = \emptyset, 1 \leq j \leq m \end{cases}$$

$$\Theta^+(X_2) = \begin{bmatrix} \theta_{11}^+ & \theta_{12}^+ & \cdots & \theta_{1m}^+ \\ \theta_{21}^+ & \theta_{22}^+ & \cdots & \theta_{2m}^+ \\ \vdots & \vdots & \theta_{ij}^+ & \vdots \\ \theta_{n1}^+ & \theta_{n2}^+ & \cdots & \theta_{nm}^+ \\ \theta_{(n+1)1}^+ & \theta_{(n+1)2}^+ & \cdots & \theta_{(n+1)m}^+ \end{bmatrix} = (\theta_{ij}^+)_{(n+1) \times m}$$

其中

$$\begin{cases} \theta_{ij}^+ = \theta_{ij}, & 1 \leq i \leq n, 1 \leq j \leq m \\ \theta_{(n+1)j}^+ = 1, & \delta_{A_j}(x_{n+1}) \subseteq X_2, 1 \leq j \leq m \\ \theta_{(n+1)j}^+ = 0, & \delta_{A_j}(x_{n+1}) \not\subseteq X_2, 1 \leq j \leq m \end{cases}$$

$$\Omega^+(X_2) = \begin{bmatrix} \omega_{11}^+ & \omega_{12}^+ & \cdots & \omega_{1m}^+ \\ \omega_{21}^+ & \omega_{22}^+ & \cdots & \omega_{2m}^+ \\ \vdots & \vdots & \omega_{ij}^+ & \vdots \\ \omega_{n1}^+ & \omega_{n2}^+ & \cdots & \omega_{nm}^+ \\ \omega_{(n+1)1}^+ & \omega_{(n+1)2}^+ & \cdots & \omega_{(n+1)m}^+ \end{bmatrix} = (\omega_{ij}^+)_{(n+1) \times m}$$

其中

$$\begin{cases} \omega_{ij}^+ = 1, & \omega_{ij} = 1, 1 \leq i \leq n, 1 \leq j \leq m \\ \omega_{ij}^+ = 1, & \omega_{ij} = 0 \wedge x_i \in \delta_{A_j}(x_{n+1}), 1 \leq i \leq n, 1 \leq j \leq m \\ \omega_{ij}^+ = 0, & \omega_{ij} = 0 \wedge x_i \notin \delta_{A_j}(x_{n+1}), 1 \leq i \leq n, 1 \leq j \leq m \\ \omega_{(n+1)j}^+ = 1, & 1 \leq j \leq m \end{cases}$$

证明 1) 对于信息系统新增的对象 x_{n+1} , 在属性子集 A_j 下的邻域类为 $\delta_{A_j}(x_{n+1})$, 根据邻域关系的对称性, 若 $x_i \in \delta_{A_j}(x_{n+1})$, 那么必有 $x_{n+1} \in \delta_{A_j}(x_i)$, 由于 $x_{n+1} \notin X_1$, 所以 $\delta_{A_j}(x_i) \not\subseteq X_1$, 即当 $x_i \in \delta_{A_j}(x_{n+1})$ 时满足 $\theta_{ij}^+ = 0 (1 \leq i \leq n, 1 \leq j \leq m)$, 同理若 $x_i \notin \delta_{A_j}(x_{n+1})$ 则 $x_{n+1} \notin \delta_{A_j}(x_i)$, 因此当 $x_i \notin \delta_{A_j}(x_{n+1})$ 时满足 $\theta_{ij}^+ = \theta_{ij} (1 \leq i \leq n, 1 \leq j \leq m)$ 。对于新增的对象 x_{n+1} , 必满足关系

$x_{n+1} \in \delta_{A_j}(x_{n+1})$, 而 $x_{n+1} \notin X_1$, 则 $\delta_{A_j}(x_{n+1}) \not\subseteq X_1$, 所以有 $\theta_{(n+1)j}^+ = 0 (1 \leq j \leq m)$ 。

2) 对于对象 $\forall x_i \in U$, $\delta_{A_j}(x_i)$ 表示对象 x_i 在论域 U 下的邻域类, 由于 $x_{n+1} \notin X_1$, 若 $\delta_{A_j}(x_i) \cap X \neq \emptyset$, 则 $(\delta_{A_j}(x_i) \cup \{x_{n+1}\}) \cap X_1 \neq \emptyset$, 若 $\delta_{A_j}(x_i) \cap X = \emptyset$, 则 $(\delta_{A_j}(x_i) \cup \{x_{n+1}\}) \cap X_1 = \emptyset$, 因此满足 $\omega_{ij}^+ = \omega_{ij} (1 \leq i \leq n, 1 \leq j \leq m)$ 。同时若 $\delta_{A_j}(x_{n+1}) \cap X_1 \neq \emptyset (1 \leq j \leq m)$, 则 $\omega_{(n+1)j}^+ = 1$, 若 $\delta_{A_j}(x_{n+1}) \cap X_1 = \emptyset (1 \leq j \leq m)$, 则 $\omega_{(n+1)j}^+ = 0$ 。

3) 对于对象 $\forall x_i \in U$, $\delta_{A_j}(x_i)$ 表示对象 x_i 在论域 U 下的邻域类, 由于 $X_2 = X \cup \{x_{n+1}\}$, 若 $\delta_{A_j}(x_i) \subseteq X$, 那么 $(\delta_{A_j}(x_i) \cup \{x_{n+1}\}) \subseteq (X \cup \{x_{n+1}\})$, 即为 $(\delta_{A_j}(x_i) \cup \{x_{n+1}\}) \subseteq X_2$; 若 $\delta_{A_j}(x_i) \not\subseteq X$, 那么 $(\delta_{A_j}(x_i) \cup \{x_{n+1}\}) \not\subseteq X_2$, 因此 $\theta_{ij}^+ = \theta_{ij} (1 \leq i \leq n, 1 \leq j \leq m)$ 。同时, 若 $\delta_{A_j}(x_{n+1}) \subseteq X_2 (1 \leq j \leq m)$, 那么 $\theta_{(n+1)j}^+ = 1$, 若 $\delta_{A_j}(x_{n+1}) \not\subseteq X_2 (1 \leq j \leq m)$, 那么 $\theta_{(n+1)j}^+ = 0$ 。

4) 对于对象 $\forall x_i \in U$, $\delta_{A_j}(x_i)$ 表示对象 x_i 在论域 U 下的邻域类, 若 $\delta_{A_j}(x_i) \cap X \neq \emptyset$, 那么 $\delta_{A_j}(x_i) \cap (X \cup \{x_{n+1}\}) \neq \emptyset$, 由于 $X_2 = X \cup \{x_{n+1}\}$, 所以当 $\omega_{ij} = 1$ 时有 $\omega_{ij}^+ = 1 (1 \leq i \leq n, 1 \leq j \leq m)$ 。若 $\omega_{ij} = 0$, 则 $\delta_{A_j}(x_i) \cap X = \emptyset$, 如果 $x_i \in \delta_{A_j}(x_{n+1})$, 根据邻域关系的对称性, 那么对象 x_i 在论域 U^+ 下的邻域类 $\delta'_{A_j}(x_i)$ 包含对象 x_{n+1} , 因此 $\delta'_{A_j}(x_i) \cap X_2 \neq \emptyset$, 即 $\omega_{ij} = 0 \wedge x_i \in \delta_{A_j}(x_{n+1})$ 满足 $\omega_{ij}^+ = 1 (1 \leq i \leq n, 1 \leq j \leq m)$, 同理当 $\omega_{ij} = 0 \wedge x_i \notin \delta_{A_j}(x_{n+1})$ 时满足 $\omega_{ij}^+ = 0 (1 \leq i \leq n, 1 \leq j \leq m)$ 。此外, 必有 $x_{n+1} \in \delta_{A_j}(x_{n+1})$, 因此 $\delta_{A_j}(x_{n+1}) \cap X_2 \neq \emptyset$, 所以 $\omega_{ij}^+ = 1 (1 \leq j \leq m)$ 。

性质 3 和性质 4 表明了, 在邻域多粒度粗糙集模型中, 所提出的子集和交集近似关系矩阵在上下近似集的计算中发挥了关键的作用, 并且可以直接通过两种近似关系矩阵计算出邻域多粒度粗糙集的上下近似集结果。因此, 当信息系统的论域增加单个对象时, 定理 1 通过增量式的方法重新地计算了两种近似关系矩阵, 并且该增量式更新方法仅对新增的对象计算了邻域类, 极大地提高了计算效率。得到了新信息系统下的子集和交集近似关系矩阵, 借助于性质 3 和性质 4 便可以直接得到新信息系统下的邻域多粒度粗糙集模型结果。

定理 1 给出了当信息系统增加单个对象时, 两种近似关系矩阵的增量式计算, 当信息系统批量地增加多个对象时, 那么可以将这多个对象分解成单个对象进行逐个增加, 然后逐步采用定理 1 进行迭代计算, 直至完成最终近似关系矩阵的增量式更新。

定理 2 给定数值型信息系统 $S = (U, A_T)$, $U = \{x_1, x_2, \dots, x_n\}$, 邻域半径为 δ 。近似目标对象集 X 在属性子集族 A_1, A_2, \dots, A_m 下的子集近似关系矩阵和

交集近似关系矩阵分别为 $\Theta(X)$ 和 $\Omega(X)$ 。当数值型信息系统论域增加对象集 $\{x_{n+1}, x_{n+2}, \dots, x_{n+s}\}$ 时, 新的论域为 $U^+ = U \cup \{x_{n+1}, x_{n+2}, \dots, x_{n+s}\}$, 新的近似目标对象集为 X^+ , 那么论域 U^+ 下 X^+ 在属性子集族 A_1, A_2, \dots, A_m 下的子集近似关系矩阵 $\Theta^+(X^+)$ 和交集近似关系矩阵 $\Omega^+(X^+)$ 分别增量式更新为

$$\Theta^+(X^+) = \Theta^{(s)}(X^{(s)}) = f_1(\Theta^{(s-1)}(X^{(s-1)}))$$

$$\Omega^+(X^+) = \Omega^{(s)}(X^{(s)}) = f_1(\Omega^{(s-1)}(X^{(s-1)}))$$

这里的 $X^{(i)} (1 \leq i \leq s)$ 表示信息系统论域 U 增加了对象集 $\{x_{n+1}, x_{n+2}, \dots, x_{n+i}\}$ 后, 近似目标对象集的最新结果, 并且 $X = X^{(0)}$ 。 $\Theta^{(i)}(X^{(i)})$ 和 $\Omega^{(i)}(X^{(i)}) (1 \leq i \leq s)$ 分别表示信息系统论域 U 增加了对象集 $\{x_{n+1}, x_{n+2}, \dots, x_{n+i}\}$ 后近似关系矩阵的更新结果, 并且 $\Theta(X) = \Theta^{(0)}(X^{(0)})$, $\Omega(X) = \Omega^{(0)}(X^{(0)})$ 。 $f_1(\cdot): \mathfrak{R}^{(n+i-1) \times m} \rightarrow \mathfrak{R}^{(n+i) \times m}$ 为定理 1 所示的近似关系矩阵 $\Theta^{(i-1)}(X^{(i-1)})$ 至 $\Theta^{(i)}(X^{(i)})$ ($\Omega^{(i-1)}(X^{(i-1)})$ 至 $\Omega^{(i)}(X^{(i)})$) 的更新映射关系函数。

基于定理 1 所示的单个对象更新映射关系 $f_1(\cdot)$, 通过定理 2 进行迭代计算, 我们可以快速高效地完成信息系统增加多个对象时两种近似关系矩阵的增量式更新, 利用最新的近似关系矩阵结果, 结合性质 3 和性质 4 便可以得到最终的邻域多粒度粗糙集的上下近似集, 在计算的过程中, 仅对增加的对象集 $\{x_{n+1}, x_{n+2}, \dots, x_{n+s}\}$ 计算了每个对象的邻域类 $\delta(x_{n+i}), 1 \leq i \leq s$, 而没有对原先论域 U 中的对象进行邻域类的重复计算, 因此该更新计算方法具有较高的更新计算效率。

2.3 对象减少时邻域多粒度粗糙集的增量式更新

2.2 节针对信息系统对象增加的情形提出了邻域多粒度粗糙集模型的增量式更新计算方法, 本节采用类似的研究思路, 提出信息系统对象减少时邻域多粒度粗糙集的增量式更新计算方法。

定理 3 给定数值型信息系统 $S = (U, A_T)$, $U = \{x_1, x_2, \dots, x_n\}$, 设属性子集族 A_1, A_2, \dots, A_m , 邻域半径为 δ 。近似目标对象集 $X \subseteq U$ 在 A_1, A_2, \dots, A_m 下的子集近似关系矩阵和交集近似关系矩阵分别为 $\Theta(X) = (\theta_{ij})_{n \times m}$ 和 $\Omega(X) = (\omega_{ij})_{n \times m}$ 。当数值型信息系统论域减少对象 $x_i, x_i \in U, U^- = U - \{x_i\}$, 对象 x_i 在属性子集 $A_i (1 \leq i \leq m)$ 下关于论域 U 的邻域类为 $\delta_{A_i}(x_i)$ 。对于两种新的近似目标对象集 $X_1 = X$ 和 $X_2 = X - \{x_i\}$, 论域 U^- 下 X_1 在 A_1, A_2, \dots, A_m 下的子集近似关系矩阵和交集近似关系矩阵分别为 $\Theta^-(X_1)$ 和 $\Omega^-(X_1)$, 论域 U^- 下 X_2 在 A_1, A_2, \dots, A_m 下的子集近似关系矩阵和交集近似关系矩阵分别为 $\Theta^-(X_2)$ 和 $\Omega^-(X_2)$, 那么它们可增量式计算为

$$\theta^-(X_1) = \begin{bmatrix} \theta_{11}^- & \theta_{12}^- & \cdots & \theta_{1m}^- \\ \theta_{21}^- & \theta_{22}^- & \cdots & \theta_{2m}^- \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{(t-1)1}^- & \theta_{(t-1)2}^- & \cdots & \theta_{(t-1)m}^- \\ \text{null} & \text{null} & \cdots & \text{null} \\ \theta_{(t+1)1}^- & \theta_{(t+1)2}^- & \cdots & \theta_{(t+1)m}^- \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{n1}^- & \theta_{n2}^- & \cdots & \theta_{nm}^- \end{bmatrix} = (\theta_{ij}^-)_{n \times m}$$

式中

$$\begin{cases} \theta_{ij}^- = 0, & \theta_{ij} = 0 \wedge x_i \in \delta_{A_j}(x_t) \wedge (\delta_{A_j}(x_i) - \{x_t\}) \not\subseteq X_1, \\ & 1 \leq i \neq t \leq n, 1 \leq j \leq m \\ \theta_{ij}^- = 1, & \theta_{ij} = 0 \wedge x_i \in \delta_{A_j}(x_t) \wedge (\delta_{A_j}(x_i) - \{x_t\}) \subseteq X_1, \\ & 1 \leq i \neq t \leq n, 1 \leq j \leq m \\ \theta_{ij}^- = 1, & \theta_{ij} = 1, 1 \leq i \neq t \leq n, 1 \leq j \leq m \\ \theta_{ij}^- = \text{null}, & 1 \leq j \leq m \end{cases}$$

其中 null 表示数值已被删除。

$$\omega^-(X_1) = \begin{bmatrix} \omega_{11}^- & \omega_{12}^- & \cdots & \omega_{1m}^- \\ \omega_{21}^- & \omega_{22}^- & \cdots & \omega_{2m}^- \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{(t-1)1}^- & \omega_{(t-1)2}^- & \cdots & \omega_{(t-1)m}^- \\ \text{null} & \text{null} & \cdots & \text{null} \\ \omega_{(t+1)1}^- & \omega_{(t+1)2}^- & \cdots & \omega_{(t+1)m}^- \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{n1}^- & \omega_{n2}^- & \cdots & \omega_{nm}^- \end{bmatrix} = (\omega_{ij}^-)_{n \times m}$$

其中

$$\begin{cases} \omega_{ij}^- = \omega_{ij}, & 1 \leq i \neq t \leq n, 1 \leq j \leq m \\ \omega_{ij}^- = \text{null}, & 1 \leq j \leq m \end{cases}$$

$$\theta^-(X_2) = \begin{bmatrix} \theta_{11}^- & \theta_{12}^- & \cdots & \theta_{1m}^- \\ \theta_{21}^- & \theta_{22}^- & \cdots & \theta_{2m}^- \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{(t-1)1}^- & \theta_{(t-1)2}^- & \cdots & \theta_{(t-1)m}^- \\ \text{null} & \text{null} & \cdots & \text{null} \\ \theta_{(t+1)1}^- & \theta_{(t+1)2}^- & \cdots & \theta_{(t+1)m}^- \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{n1}^- & \theta_{n2}^- & \cdots & \theta_{nm}^- \end{bmatrix} = (\theta_{ij}^-)_{n \times m}$$

其中

$$\begin{cases} \theta_{ij}^- = \theta_{ij}, & 1 \leq i \neq t \leq n, 1 \leq j \leq m \\ \theta_{ij}^- = \text{null}, & 1 \leq j \leq m \end{cases}$$

$$\omega^-(X_2) = \begin{bmatrix} \omega_{11}^- & \omega_{12}^- & \cdots & \omega_{1m}^- \\ \omega_{21}^- & \omega_{22}^- & \cdots & \omega_{2m}^- \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{(t-1)1}^- & \omega_{(t-1)2}^- & \cdots & \omega_{(t-1)m}^- \\ \text{null} & \text{null} & \cdots & \text{null} \\ \omega_{(t+1)1}^- & \omega_{(t+1)2}^- & \cdots & \omega_{(t+1)m}^- \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{n1}^- & \omega_{n2}^- & \cdots & \omega_{nm}^- \end{bmatrix} = (\omega_{ij}^-)_{n \times m}$$

其中

$$\begin{cases} \omega_{ij}^- = 0, & \omega_{ij} = 1 \wedge x_i \in \delta_{A_j}(x_t) \wedge (\delta_{A_j}(x_i) - \{x_t\}) \cap X_2 = \emptyset, \\ & 1 \leq i \neq t \leq n, 1 \leq j \leq m \\ \omega_{ij}^- = 1, & \omega_{ij} = 1 \wedge x_i \in \delta_{A_j}(x_t) \wedge (\delta_{A_j}(x_i) - \{x_t\}) \cap X_2 \neq \emptyset, \\ & 1 \leq i \neq t \leq n, 1 \leq j \leq m \\ \omega_{ij}^- = \omega_{ij}, & x_i \notin \delta_{A_j}(x_t), 1 \leq i \neq t \leq n, 1 \leq j \leq m \\ \omega_{ij}^- = \text{null}, & 1 \leq j \leq m \end{cases}$$

证明 1) 根据邻域关系的对称性, 对于 $x_t \in U$, 若 $x_i \in \delta_{A_j}(x_t)$, 那么 $x_t \in \delta_{A_j}(x_i)$, 那么如果 $\theta_{ij} = 0 \wedge (\delta_{A_j}(x_i) - \{x_t\}) \not\subseteq X_1$, 则满足关系 $\theta_{ij}^- = 0 (1 \leq i \neq t \leq n, 1 \leq j \leq m)$ 。如果 $\theta_{ij} = 0 \wedge (\delta_{A_j}(x_i) - \{x_t\}) \subseteq X_1$, 则 $\theta_{ij}^- = 1 (1 \leq i \neq t \leq n, 1 \leq j \leq m)$ 。对于 $\theta_{ij} = 1$, 那么 $\delta_{A_j}(x_i) \subseteq X$, 即 $x_t \notin \delta_{A_j}(x_i)$, 所以 $\delta_{A_j}(x_i) - x_t = \delta_{A_j}(x_i)$, 则 $\delta_{A_j}(x_i) \subseteq X_1$, 因此 $\theta_{ij} = 1$ 时 $\theta_{ij}^- = 1 (1 \leq i \neq t \leq n, 1 \leq j \leq m)$ 。由于 x_t 已被移除, 因此 $\theta_{ij}^- = \text{null} (1 \leq j \leq m)$ 。

2) 对于 $x_i \in U (1 \leq i \neq t \leq n)$, 由于 $x_t \notin X, X = X_1$, 因此 $(\delta_{A_j}(x_i) - \{x_t\}) \cap X_1 = \delta_{A_j}(x_i) \cap X, 1 \leq j \leq m$ 。那么有 $\omega_{ij}^- = \omega_{ij}, (1 \leq i \neq t \leq n, 1 \leq j \leq m), \omega_{ij}^- = \text{null}, 1 \leq j \leq m$ 。

3) 对于 $x_i \in U (1 \leq i \neq t \leq n)$, 由于 $x_t \in X, X - \{x_t\} = X_2$, 因此满足关系 $(\delta_{A_j}(x_i) - \{x_t\}) \cap (X - \{x_t\}) = \delta_{A_j}(x_i) \cap X, 1 \leq j \leq m$ 。那么有 $\theta_{ij}^- = \theta_{ij}, (1 \leq i \neq t \leq n, 1 \leq j \leq m), \theta_{ij}^- = \text{null}, 1 \leq j \leq m$ 。

4) 对于 $x_i \in \delta_{A_j}(x_t), (1 \leq i \neq t \leq n, 1 \leq j \leq m)$, 则 $x_t \in \delta_{A_j}(x_i)$, 如果 $\omega_{ij} = 1 \wedge (\delta_{A_j}(x_i) - \{x_t\}) \cap X_2 = \emptyset$, 那么 $\omega_{ij}^- = 0$, 同理如果 $\omega_{ij} = 1 \wedge (\delta_{A_j}(x_i) - \{x_t\}) \cap X_2 \neq \emptyset$, 那么 $\omega_{ij}^- = 1$ 。对于 $x_i \notin \delta_{A_j}(x_t)$, 那么 $\delta_{A_j}(x_i) - \{x_t\} = \delta_{A_j}(x_i)$, 那么满足 $(\delta_{A_j}(x_i) - \{x_t\}) \cap X_2 = \delta_{A_j}(x_i) \cap X_2 = \delta_{A_j}(x_i) \cap X$, 因此 $\omega_{ij}^- = \omega_{ij}$, 即 $x_i \notin \delta_{A_j}(x_t)$ 满足 $\omega_{ij}^- = \omega_{ij}$ 。

当信息系统的论域减少单个对象时, 定理 3 通过增量式的方法重新计算子集近似关系矩阵和交集近似关系矩阵, 其中仅对部分对象计算了其邻域类, 极大地提高了计算效率, 借助于性质 3 和性质 4 便可以直接得到新信息系统下的邻域多粒度粗糙集模型结果, 因此同样可以进一步提高邻域多粒度粗糙集上下近似集的更新计算效率。

类似于定理 2 对单个对象进行迭代的方法, 当信息系统批量地减少多个对象时, 那么可以将这多个对象分解成单个对象进行逐个减少, 然后逐步采用定理 3 进行迭代计算, 直至完成最终近似关系矩阵的增量式更新, 具体如定理 4 所示。

定理 4 给定数值型信息系统 $S = (U, A_T), U = \{x_1, x_2, \dots, x_n\}$, 邻域半径为 δ 。近似目标对象集 X 在属性子集族 A_1, A_2, \dots, A_m 下的子集近似关系矩阵和交集近似关系矩阵分别为 $\theta(X)$ 和 $\omega(X)$ 。当数值型信息系统论域减少对象集 $\{x_{t1}, x_{t2}, \dots, x_{ts}\}$ 时, 新的论域为 $U^- = U - \{x_{t1}, x_{t2}, \dots, x_{ts}\}$, 新的近似目标对象集为 X^- , 那么论域 U^- 下 X^- 在属性子集族 A_1, A_2, \dots ,

A_m 下的子集近似关系矩阵 $\theta^-(X^-)$ 和交集近似关系矩阵 $\Omega^-(X^-)$ 分别增量式更新为

$$\theta^-(X^-) = \theta^{(s)}(X^{(s)}) = f_2(\theta^{(s-1)}(X^{(s-1)}))$$

$$\Omega^-(X^-) = \Omega^{(s)}(X^{(s)}) = f_2(\Omega^{(s-1)}(X^{(s-1)}))$$

式中: $X^{(i)} (1 \leq i \leq s)$ 表示信息系统论域 U 减少了对象集 $\{x_{i1}, x_{i2}, \dots, x_{in}\}$ 后, 近似目标对象集的最新结果, 并且 $X = X^{(0)}$; $\theta^{(i)}(X^{(i)})$ 和 $\Omega^{(i)}(X^{(i)}) (1 \leq i \leq s)$ 分别表示信息系统论域 U 减少了对象集 $\{x_{i1}, x_{i2}, \dots, x_{in}\}$ 后近似关系矩阵的更新结果, 并且 $\theta(X) = \theta^{(0)}(X^{(0)})$, $\Omega(X) = \Omega^{(0)}(X^{(0)})$; $f_2(\cdot): \mathfrak{R}^{n \times m} \rightarrow \mathfrak{R}^{n \times m}$ 为定理 3 所示的近似关系矩阵 $\theta^{(i-1)}(X^{(i-1)})$ 至 $\theta^{(i)}(X^{(i)})$ ($\Omega^{(i-1)}(X^{(i-1)})$ 至 $\Omega^{(i)}(X^{(i)})$) 的更新映射关系函数。

基于定理 3 单个对象的更新映射关系 $f_2(\cdot)$, 通过定理 4 进行迭代计算, 我们可以快速高效地完成信息系统减少多个对象时两种近似关系矩阵的增量式更新, 利用最新的近似关系矩阵结果, 结合性质 3 和性质 4 便可以得到最终的邻域多粒度粗糙集的上下近似集, 在计算的过程中, 仅对减少的对象集 $\{x_{i1}, x_{i2}, \dots, x_{in}\}$ 计算了每个对象的邻域类 $\delta(x_{ii})$, $1 \leq i \leq s$, 而没有对原先论域 U 中的对象进行邻域类的重复计算, 因此该更新计算方法同样具有较高的更新计算效率。

3 邻域多粒度粗糙集动态更新算法

3.1 非增量式更新算法

针对信息系统对象增加和减少的情形, 本小节基于传统的邻域多粒度粗糙集计算方法, 提出对应的计算算法, 具体算法 1 和算法 2 所示, 由于该算法不具有增量式更新计算的特性, 因此该算法也被称为非增量式更新算法。

算法 1 对象增加时邻域多粒度粗糙集非增量式更新算法。

输入 数值型信息系统 $S = (U, A_T)$, $U = \{x_1, x_2, \dots, x_n\}$, 邻域半径为 δ , 属性集 A_T 确定的属性子集族 A_1, A_2, \dots, A_m , 近似目标对象集 X 。信息系统论域增加对象集 $\{x_{n+1}, x_{n+2}, \dots, x_{n+s}\}$, 新的论域 $U^+ = U \cup \{x_{n+1}, x_{n+2}, \dots, x_{n+s}\}$, 新的近似目标对象集为 X^+ 。

输出 近似目标对象集 X^+ 在属性子集族 A_1, A_2, \dots, A_m 下的乐观 (悲观) 邻域多粒度下近似集 $\varphi(X^+)$ 和上近似集 $\phi(X^+)$ 。

1) 初始化近似集 $\varphi(X^+) = \emptyset$, $\phi(X^+) = \emptyset$, 初始化两种近似关系矩阵 $\theta^+(X^+) = (\theta_{ij}^+)_{(n+s) \times m} = \mathbf{0}$ 和 $\Omega^+(X^+) = (\omega_{ij}^+)_{(n+s) \times m} = \mathbf{0}$ 。/*这里的 $\mathbf{0}$ 代表全 0 矩阵*/

2) 对于 $x_i \in U^+ (1 \leq i \leq n+s)$, 计算 x_i 在属性集 $A_j (1 \leq j \leq m)$ 下的 δ 邻域类 $\delta_{A_j}(x_i)$ 。

3) 对于 $\delta_{A_j}(x_i)$, $(1 \leq i \leq n+s, 1 \leq j \leq m)$, 若 $\delta_{A_j}(x_i) \subseteq$

X^+ , 那么 $\theta_{ij}^+ = 1$, 否则 $\theta_{ij}^+ = 0$, 若 $\delta_{A_j}(x_i) \cap X^+ \neq \emptyset$, 那么 $\omega_{ij}^+ = 1$, 否则 $\omega_{ij}^+ = 0$ 。

4) 根据两种近似关系矩阵 $\theta^+(X^+)$ 和 $\Omega^+(X^+)$ 计算 X^+ 在属性子集族 A_1, A_2, \dots, A_m 下的乐观 (悲观) 邻域多粒度下近似集 $\varphi(X^+)$ 和上近似集 $\phi(X^+)$ 。/*根据性质 3 和性质 4*/

5) 返回 $\varphi(X^+)$ 和 $\phi(X^+)$ 。

在算法 1 中, 设信息系统条件属性集的属性数量为 a , 那么整个算法 1 的计算量主要集中在所有对象邻域类的计算上, 因此整个算法 1 的时间复杂度为 $O(a \cdot (n+s)^2)$ 。

算法 2 对象减少时邻域多粒度粗糙集非增量式更新算法。

输入 数值型信息系统 $S = (U, A_T)$, $U = \{x_1, x_2, \dots, x_n\}$, 邻域半径为 δ , 属性集 A_T 确定的属性子集族 A_1, A_2, \dots, A_m , 近似目标对象集 X 。信息系统论域减少对象集 $\{x_{i1}, x_{i2}, \dots, x_{is}\}$, 新的论域 $U^- = U - \{x_{i1}, x_{i2}, \dots, x_{is}\}$, 新的近似目标对象集为 X^- 。

输出 近似目标对象集 X^- 在属性子集族 A_1, A_2, \dots, A_m 下的乐观 (悲观) 邻域多粒度下近似集 $\varphi(X^-)$ 和上近似集 $\phi(X^-)$ 。

1) 初始化近似集 $\varphi(X^-) = \emptyset$, $\phi(X^-) = \emptyset$, 初始化两种近似关系矩阵 $\theta^-(X^-) = (\theta_{ij}^-)_{(n-s) \times m} = \mathbf{0}$ 和 $\Omega^-(X^-) = (\omega_{ij}^-)_{(n-s) \times m} = \mathbf{0}$ 。/*这里的 $\mathbf{0}$ 代表全 0 矩阵*/

2) 对于 $x_i \in U^- (1 \leq i \leq n-s)$, 计算 x_i 在属性集 $A_j (1 \leq j \leq m)$ 下的 δ 邻域类 $\delta_{A_j}(x_i)$ 。

3) 对于 $\delta_{A_j}(x_i)$, $(1 \leq i \leq n-s, 1 \leq j \leq m)$, 若 $\delta_{A_j}(x_i) \subseteq X^-$, 那么 $\theta_{ij}^- = 1$, 否则 $\theta_{ij}^- = 0$, 若 $\delta_{A_j}(x_i) \cap X^- \neq \emptyset$, 那么 $\omega_{ij}^- = 1$, 否则 $\omega_{ij}^- = 0$ 。

4) 根据两种近似关系矩阵 $\theta^-(X^-)$ 和 $\Omega^-(X^-)$ 计算 X^- 在属性子集族 A_1, A_2, \dots, A_m 下的乐观 (悲观) 邻域多粒度下近似集 $\varphi(X^-)$ 和上近似集 $\phi(X^-)$ 。/*根据性质 3 和性质 4*/

5) 返回 $\varphi(X^-)$ 和 $\phi(X^-)$ 。

类似于算法 1 的时间复杂度分析, 整个算法 2 的时间复杂度为 $O(a \cdot (n-s)^2)$ 。

3.2 增量式更新算法

根据本文 2.2 节和 2.3 节所示的邻域多粒度粗糙集增量式更新方法, 接下来提出相应的增量式更新算法。

算法 3 对象增加时邻域多粒度粗糙集增量式更新算法。

输入 1) 数值型信息系统 $S = (U, A_T)$, $U = \{x_1, x_2, \dots, x_n\}$, 邻域半径为 δ , 属性集 A_T 确定的属性子集族 A_1, A_2, \dots, A_m , 近似目标对象集 X , X 在属性子集族 A_1, A_2, \dots, A_m 下的两种近似关系矩阵分别为 $\theta(X)$

和 $\Omega(X)$ 。

2) 论域增加的对象集为 $\{x_{n+1}, x_{n+2}, \dots, x_{n+s}\}$, 新的论域表示为 $U^+ = U \cup \{x_{n+1}, x_{n+2}, \dots, x_{n+s}\}$, 新的近似目标对象集为 X^+ 。

输出 近似目标对象集 X^+ 在属性子集族 A_1, A_2, \dots, A_m 下的乐观(悲观)邻域多粒度下近似集 $\varphi(X^+)$ 和上近似集 $\phi(X^+)$ 。

1) 初始化近似集 $\varphi(X^+) = \emptyset, \phi(X^+) = \emptyset$, 初始化两种近似关系矩阵 $\Theta^+(X^+) = (\theta_{ij}^+)_{(n+s) \times m} = \mathbf{0}$ 和 $\Omega^+(X^+) = (\omega_{ij}^+)_{(n+s) \times m} = \mathbf{0}$ 。/*这里的 $\mathbf{0}$ 代表全 0 矩阵*/

2) 对于 $x_{n+i} \in \{x_{n+1}, x_{n+2}, \dots, x_{n+s}\} (1 \leq i \leq s)$, 记信息系统 $S^{(i)} = (U^{(i)} = U^{(i-1)} \cup \{x_{n+i}\}, A_T)$, 其中 $U^{(0)} = U$, 对应的近似目标集为 $X^{(i)}$, 其中 $X^{(0)} = X$ 。

3) 计算论域 $U^{(i)}$ 下对象 $x_{n+i} (1 \leq i \leq s)$ 在属性子集 $A_j (1 \leq j \leq m)$ 的邻域类为 $\delta_{A_j}(x_{n+i})$ 。

4) 在近似关系矩阵 $\Theta^{(i-1)}(X^{(i-1)})$ 和 $\Omega^{(i-1)}(X^{(i-1)})$ 的基础上更新计算论域 $U^{(i)}$ 下 $X^{(i)}$ 关于属性子集族 A_1, A_2, \dots, A_m 的近似关系矩阵 $\Theta^{(i)}(X^{(i)})$ 和 $\Omega^{(i)}(X^{(i)})$, 其中 $\Theta^{(0)}(X^{(0)}) = \Theta(X)$ 且 $\Omega^{(0)}(X^{(0)}) = \Omega(X)$ 。/*根据定理 1*/

5) 重复迭代 2)~4)。/*定理 2*/

6) 令近似关系矩阵 $\Theta^+(X^+) = \Theta^{(s)}(X^{(s)})$ 、 $\Omega^+(X^+) = \Omega^{(s)}(X^{(s)})$ 和 $X^+ = X^{(s)}$, 根据 $\Theta^+(X^+)$ 和 $\Omega^+(X^+)$ 计算 X^+ 在属性子集族 A_1, A_2, \dots, A_m 下的乐观(悲观)邻域多粒度下近似集 $\varphi(X^+)$ 和上近似集 $\phi(X^+)$ 。/*根据性质 3 和性质 4*/

7) 返回 $\varphi(X^+)$ 和 $\phi(X^+)$ 。

算法 3 在执行的过程中, 会不断地重复迭代 2)~4), 其中每次迭代所需的时间复杂度为 $O(a \cdot n)$, 直到新增的 s 个对象全部迭代完成, 因此整个算法 3 的时间复杂度为 $O(a \cdot s \cdot n)$ 。

算法 4 对象减少时邻域多粒度粗糙集增量式更新算法。

输入 1) 数值型信息系统 $S = (U, A_T)$, $U = \{x_1, x_2, \dots, x_n\}$, 邻域半径为 δ , 属性集 A_T 确定的属性子集族 A_1, A_2, \dots, A_m , 近似目标对象集 X , X 在属性子集族 A_1, A_2, \dots, A_m 下的两种近似关系矩阵分别为 $\Theta(X)$ 和 $\Omega(X)$ 。

2) 论域减少的对象集为 $\{x_{t1}, x_{t2}, \dots, x_{ts}\}$, 新的论域表示为 $U^- = U - \{x_{t1}, x_{t2}, \dots, x_{ts}\}$, 新的近似目标对象集为 X^- 。

输出 近似目标对象集 X^- 在属性子集族 A_1, A_2, \dots, A_m 下的乐观(悲观)邻域多粒度下近似集 $\varphi(X^-)$ 和上近似集 $\phi(X^-)$ 。

1) 初始化近似集 $\varphi(X^-) = \emptyset, \phi(X^-) = \emptyset$, 初始化两种近似关系矩阵 $\Theta^-(X^-) = (\theta_{ij}^-)_{n \times m} = \mathbf{0}$ 和 $\Omega^-(X^-) = (\omega_{ij}^-)_{n \times m} = \mathbf{0}$ 。/*这里的 $\mathbf{0}$ 代表全 0 矩阵*/

2) 对于 $x_{ti} \in \{x_{t1}, x_{t2}, \dots, x_{ts}\} (1 \leq i \leq s)$, 记信息系统 $S^{(i)} = (U^{(i)} = U^{(i-1)} - \{x_{ti}\}, A_T)$, 其中 $U^{(0)} = U$, 对应的近似目标集为 $X^{(i)}$, 其中 $X^{(0)} = X$ 。

3) 计算论域 $U^{(i)}$ 下对象 $x_{ti} (1 \leq i \leq s)$ 在属性子集 $A_j (1 \leq j \leq m)$ 的邻域类为 $\delta_{A_j}(x_{ti})$ 。

4) 在近似关系矩阵 $\Theta^{(i-1)}(X^{(i-1)})$ 和 $\Omega^{(i-1)}(X^{(i-1)})$ 的基础上更新计算论域 $U^{(i)}$ 下 $X^{(i)}$ 关于属性子集族 A_1, A_2, \dots, A_m 的近似关系矩阵 $\Theta^{(i)}(X^{(i)})$ 和 $\Omega^{(i)}(X^{(i)})$, 其中 $\Theta^{(0)}(X^{(0)}) = \Theta(X)$ 且 $\Omega^{(0)}(X^{(0)}) = \Omega(X)$ 。/*根据定理 3*/

5) 重复迭代 2)~4)。/*定理 4*/

6) 令近似关系矩阵 $\Theta^-(X^-) = \Theta^{(s)}(X^{(s)})$ 、 $\Omega^-(X^-) = \Omega^{(s)}(X^{(s)})$ 和 $X^- = X^{(s)}$, 根据 $\Theta^-(X^-)$ 和 $\Omega^-(X^-)$ 计算 X^- 在属性子集族 A_1, A_2, \dots, A_m 下的乐观(悲观)邻域多粒度下近似集 $\varphi(X^-)$ 和上近似集 $\phi(X^-)$ 。/*根据性质 3 和性质 4*/

7) 返回 $\varphi(X^-)$ 和 $\phi(X^-)$ 。

与算法 3 类似, 整个算法 4 的时间复杂度为 $O(a \cdot s \cdot n)$ 。

4 实验分析

4.1 实验设置

为了测试本文所提出的邻域多粒度粗糙集增量式更新算法的有效性, 本节将通过实验分析的方式进行验证。本实验分为两部分, 第 1 部分将文中所提出的邻域多粒度粗糙集的增量式更新算法与非增量式更新算法进行比较, 由于这两种算法计算的是同一种模型, 因此只对它们的算法效率进行比较便可以验证增量式算法的有效性, 第 2 部分是将本文提出的增量式算法与文献 [26] 提出的邻域多粒度粗糙集增量式更新算法进行实验效率对比, 用来验证本文算法的优越性。表 1 给出了进行实验比较所使用的 8 个 UCI 数据集, 这些数据集均为数值型的属性值, 因此可以直接运用于邻域多粒度粗糙集模型, 同时这 8 个数据集包含了不同的大小规模, 可以用来验证算法在不同规模数据集下的增量式更新效果。

表 1 实验数据集

序号	数据集	对象数	属性数	类数
1	Iono	351	34	2
2	Wdbc	569	31	2
3	Biodeg	1 055	41	2
4	Mess	1 151	19	3
5	Gearbox	1 603	72	3
6	Segment	2 310	19	7

续表 1

序号	数据集	对象数	属性数	类数
7	Gisette	13 500	5 000	3
8	Skin	245 057	4	3

由于本文所研究的是邻域多粒度环境下的粗糙集模型, 即模型中含有多个邻域关系, 因此需要将数据集的条件属性集构造出一组属性子集族。本实验按照原始数据集的属性顺序, 对条件属性集进行大致平均划分构造出属性子集族, 由于本实验验证的是增量式算法与非增量式算法的算法效率, 而邻域多粒度粗糙集的近似集结果并不是本实验关注的重点, 因此构造出属性子集族中的属性子集数量可自行约定, 对于表 1 中的数据集 Iono 构造了 7 个多粒度属性子集的属性子集族, 数据集 Wdbc 构造了 6 个多粒度属性子集的属性子集族, 数据集 Biodeg 构造了 8 个多粒度属性子集的属性子集族, 数据集 Mess 构造了 5 个多粒度属性子集的属性子集族, 数据集 Gearbox 构造了 9 个多粒度属性子集的属性子集族, 数据集 Segment 构造了 5 个多粒度属性子集的属性子集族, 数据集 Gisette 构造了 100 个多粒度属性子集的属性子集族, 数据集 Skin 构造了 2 个多粒度属性子集的属性子集族。表 2 给出了进行实验所使用的软硬件环境。

另一方面, 本文所提出的算法是针对动态数据环境下模型的增量式更新, 因此需要对表 1 中的数据集进行动态变化的构造, 这里参考其他学者对于该类问题的处理^[12-20], 将数据集的论域进

行平均分割成多个部分, 然后选择一个部分作为初始数据集, 并不不断地将剩余的部分数据集进行合并, 则构造出了信息系统动态增加更新的情形, 相反, 从完整的数据集开始, 不断地从中依次进行每个部分的删除, 因而构造出了信息系统动态减少更新的情形。本实验均按照这种处理方式进行实验, 并且将实验数据集分成 10 个部分。对于邻域多粒度粗糙集模型中的参数邻域半径 δ , 在文献 [9-11] 中, 学者们通过大量实验验证了邻域半径 $\delta = 0.15$ 时模型的近似效果较佳, 因此本文直接采纳其他学者的研究成果, 将邻域半径设置为 $\delta = 0.15$ 进行实验, 模型的近似目标集均由原始数据集的类别集来定义。本文所有实验结果均取自 10 次实验求取平均值来表示。

表 2 实验软硬件环境

Table 2 Experimental software and hardware environment

组件	配置与型号
操作系统	Windows 10
处理器	英特尔酷睿 i5-1035G1, 3.6 GHz
内存	DDR4 16 GB
硬盘	512 GB 固态硬盘
编程软件	MATLAB 2017b

4.2 非增量式与增量式算法比较

图 1 给出了信息系统论域依次增加单个部分数据时, 邻域多粒度粗糙集模型增量式更新算法与非增量式更新算法在各个数据集下更新模型的时间比较结果。

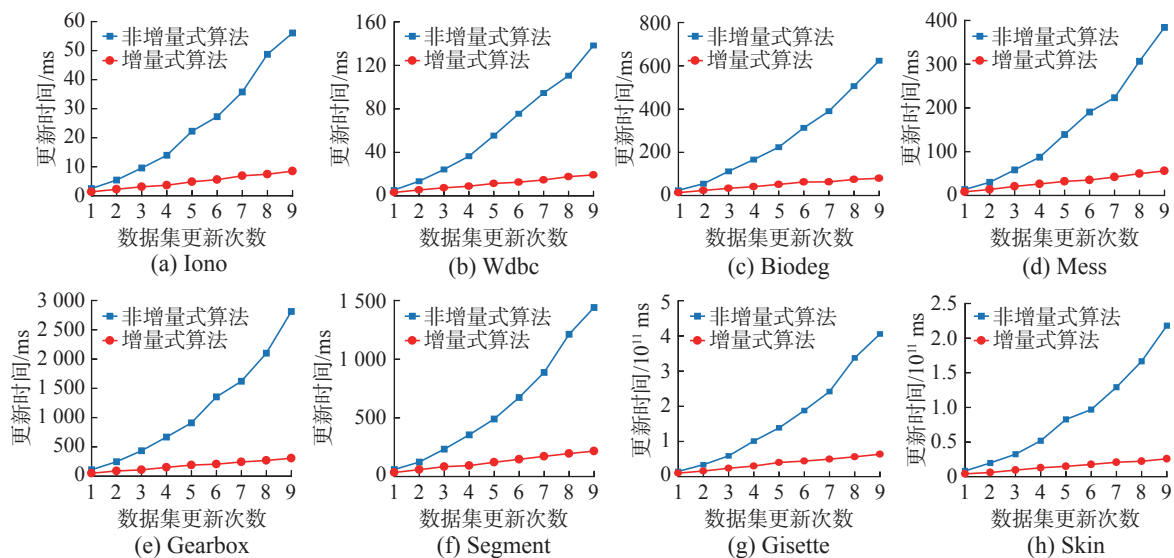


图 1 数据增加时增量式与非增量式算法更新用时比较

Fig. 1 Comparison of update time between incremental and non incremental algorithms when data increases

观察图 1 的每个子图可以发现, 随着数据集

更新次数的不断增加, 即信息系统的论域不断增

大, 2 种算法在更新邻域多粒度粗糙集模型所需的时间都是不断增加的, 但是可以很直观地看出非增量式算法的时间增长较为迅速, 增量式算法增长的非常平缓, 随着更新次数的不断增加, 2 种算法差距越来越大。例如对于数据集 Biodeg, 当完成信息系统的第 9 次更新时, 非增量式算法消耗了 620 ms, 而增量式算法仅消耗了不到 100 ms 的时间; 对于数据集 Mess, 当完成信息系统的第 9 次更新时, 非增量式算法消耗了超过 390 ms 的时间, 而增量式算法仅消耗了 50 ms。产生这一差距的主要原因是由于非增量式算法与增量式算法的计算机制不同导致的, 对于非增量式算法, 每次基于完整的数据集进行计算, 即数据集的规模越大其计算量也越大, 消耗的时间也越长, 其

消耗的时间与当前完整数据集的规模密切相关。而对于增量式算法, 通过在旧数据集的计算结果上对新增的数据进行计算, 主要的计算量集中在新增的数据上, 而对原始数据仅需要花费较少的计算, 便可以完成整个新信息系统模型的更新, 由于数据集每次新增的数据量比较少且是固定的, 因此增量式算法的更新计算时间增长得较为缓慢, 整体计算时间处于一个比较低的水平, 因此对于信息系统对象增加的情形, 本文提出的增量式更新算法具有较高的更新计算效率。

图 2 给出了信息系统论域依次减少单个部分数据时, 邻域多粒度粗糙集模型增量式更新算法与非增量式更新算法在各个数据集下更新模型的时间比较结果。

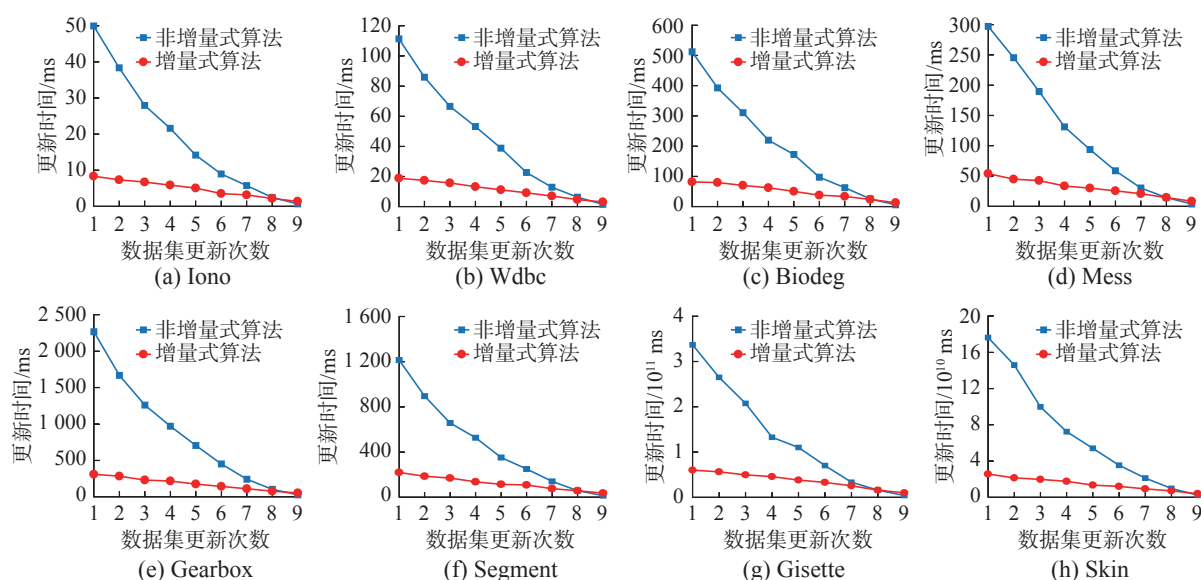


图 2 数据减少时增量式与非增量式算法更新用时比较

Fig. 2 Comparison of update time between incremental and non-incremental algorithms when data decreases

观察图 2 的每个子图, 可以发现当信息系统刚开始更新时, 非增量式算法和增量式算法的计算时间就表现出了很大的差距, 随着信息系统更新次数的增加, 两类算法的计算时间均逐渐减小, 并且差距也减小, 但是增量式算法的计算时间始终处于一个较低的水平。例如对于数据集 Wdbc, 第一次动态更新时, 非增量式算法的更新时间接近 120 ms, 而增量式算法仅消耗了 20 ms。出现这一差距同样是由于 2 种算法的计算机制不同导致的, 对于非增量式算法, 每次基于完整的数据集进行模型的计算, 而增量式算法针对变化的数据在原先旧数据集的基础上进行进一步计算, 因此增量式算法的效率大幅度高于非增量式算法。另外, 观察图 2 可以发现, 在信息系统的第 8 次和第 9 次更新时, 非增量式算法与增量式

算法的更新用时基本接近, 甚至非增量式算法的更新用时略低于增量式算法, 此时数据集经过不断更新减少, 其规模已经很小了, 因此对于小规模的数据集, 非增量式算法和增量式算法的效率相当, 但是对于大规模的数据集, 本文增量式算法的效率还是大幅度高于非增量式算法。

综合图 1 和图 2 的实验结果, 针对信息系统对象的增加和对象的减少, 所提出的邻域多粒度粗糙集增量式更新算法都具有更高的更新效率。

4.3 与同类型增量式更新算法比较

为了进一步验证所提出算法的优越性, 接下来将本文的增量式算法与文献 [26] 提出的邻域多粒度粗糙集增量式算法进行实验比较, 给定同一邻域半径和同一属性子集族, 那么这 2 种算法的模型计算结果是一致的, 因此只需对比 2 种算法

的效率便可以验证它们的性能。

图3和图4分别给出了信息系统依次增加和依次减少单个部分数据时, 2种增量式算法的更新时间比较结果。观察图3中各个数据集的实验结果可以发现, 本文算法的更新性能较高于文献[26]提出的对比算法, 当信息系统刚开始更新时, 2种算法的更新所需时间几乎相当, 随着更新次数的增加, 数据集的规模逐渐增大, 2种算法的时间消耗量逐渐表现出了差距, 文献[26]算法更新时间的增长速率要略高于本文算法, 因此最终的

更新用时较高于本文算法。观察图4中各个数据集的实验结果可以发现, 随着信息系统论域的逐渐减小, 本文算法的更新时间在大部分情形下低于文献[26]的对比算法, 例如 Iono、Wdbc、Biodeg、Mess、Segment 和 Gisette 数据集的前7次更新。对于信息系统的后两次更新, 对比算法的更新时间与本文算法比较接近, 此时数据集的规模比较小, 对比增量式算法的计算效率与本文算法相差不大。通过整体比较, 可以看出本文增量式算法的性能较高于对比增量式算法。

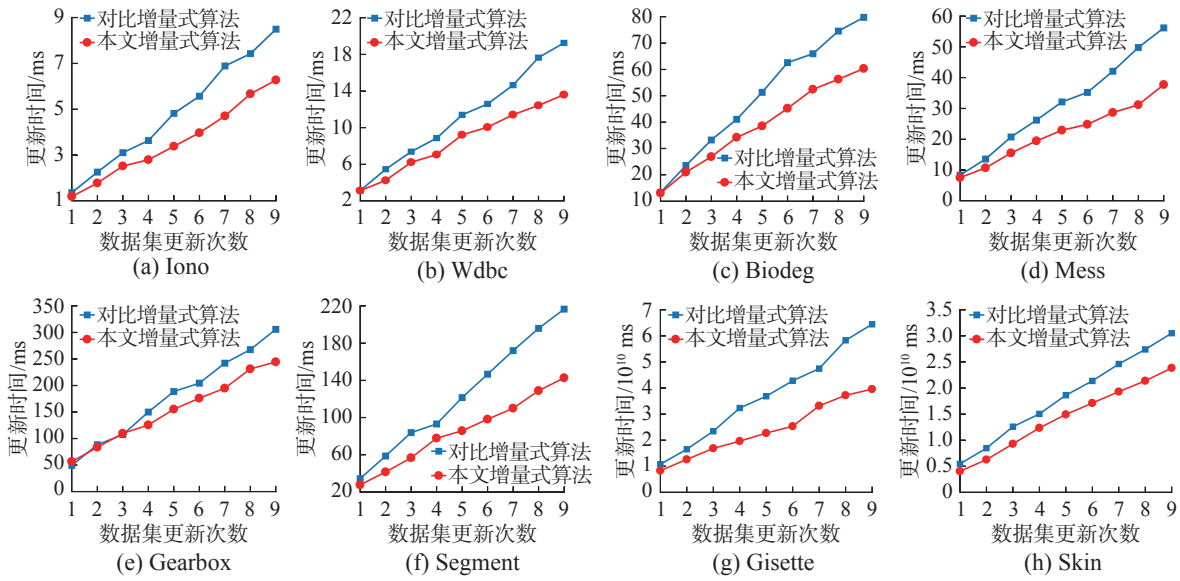


图3 数据增加时本文增量式算法与对比增量式算法更新用时比较

Fig. 3 Comparison of update time between our incremental algorithm and the compared incremental algorithm when data increases

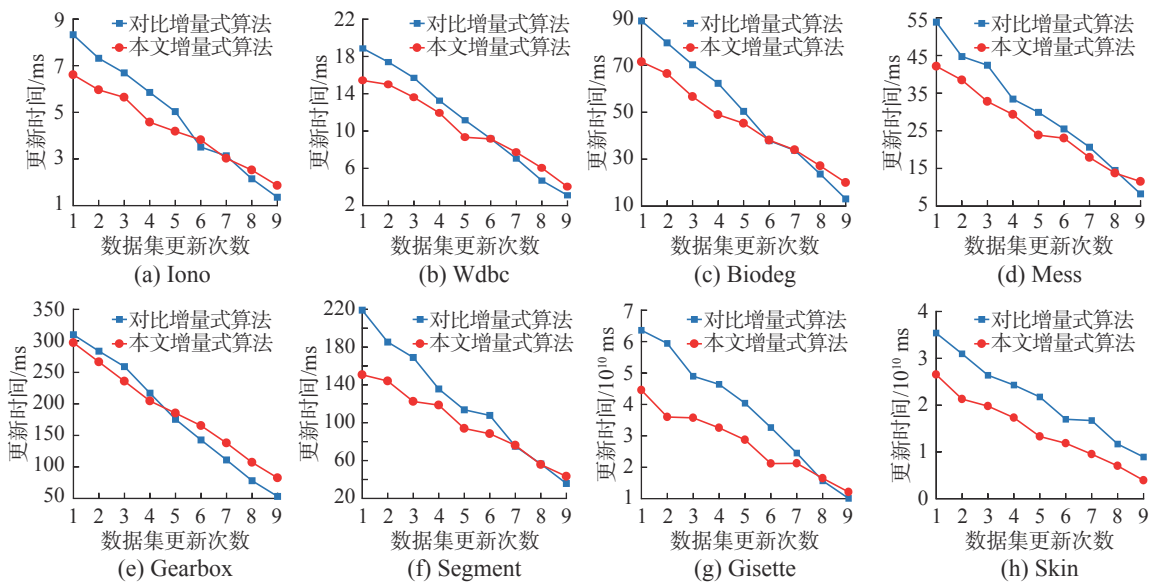


图4 数据减少时本文增量式算法与对比增量式算法更新用时比较

Fig. 4 Comparison of update time between our incremental algorithm and the compared incremental algorithm when data decreases

为了测试信息系统不同数据量变化时本文增量式算法与文献[26]对比增量式算法的更新性能,在接下来的实验中,对于信息系统论域增加的情形,首先选择信息系统的1个部分数据作为初始数据集,然后往初始数据集增加1个部分数据作为第1次增量式更新,往初始数据集一次性增加2个部分数据作为第2次增量式更新,往初始数据集一次性增加3个部分数据作为第3次增量式更新,依次类推,直到一次性增加

9个部分数据,即每次增加的数据量为信息系统整体数据量的10%、20%、…、90%。类似地,对于信息系统论域减少的情形,每次从整个信息系统开始,一次性移除1个部分数据作为第1次增量式更新,一次性移除2个部分数据作为第2次增量式更新,以此类推,直至移除9个部分数据。按照上述的数据更新,让2种算法进行邻域多粒度粗糙集的增量式更新计算,实验结果如图5、6所示。

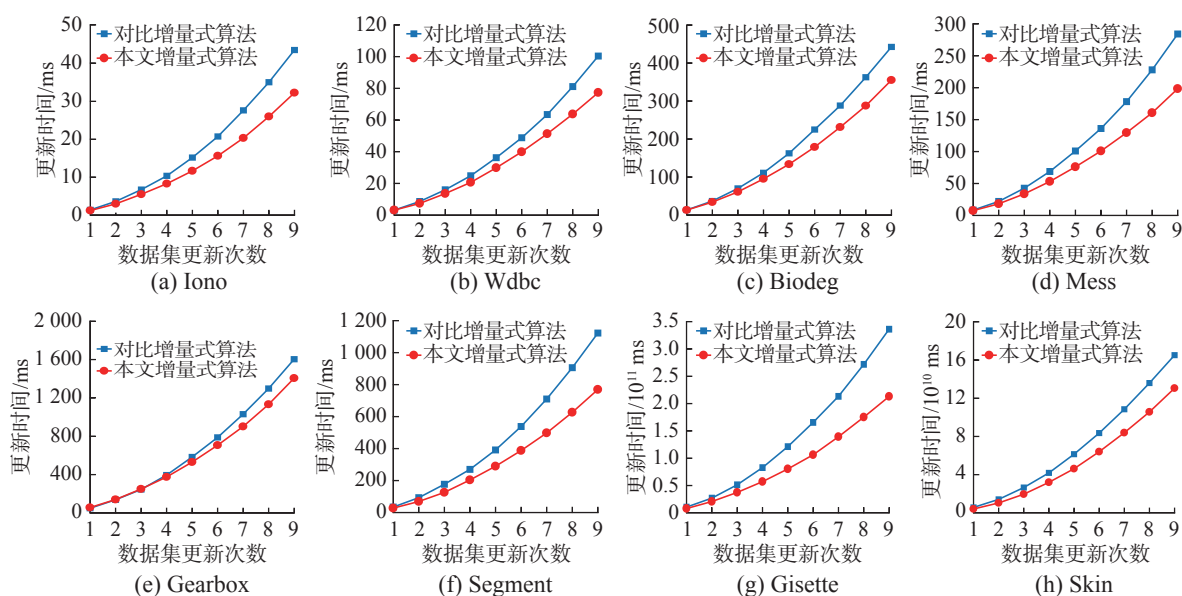


图 5 不同数据量增加时邻域多粒度粗糙集更新用时比较

Fig. 5 Comparison of updating time of neighborhood multi-granulation rough sets with different data volume

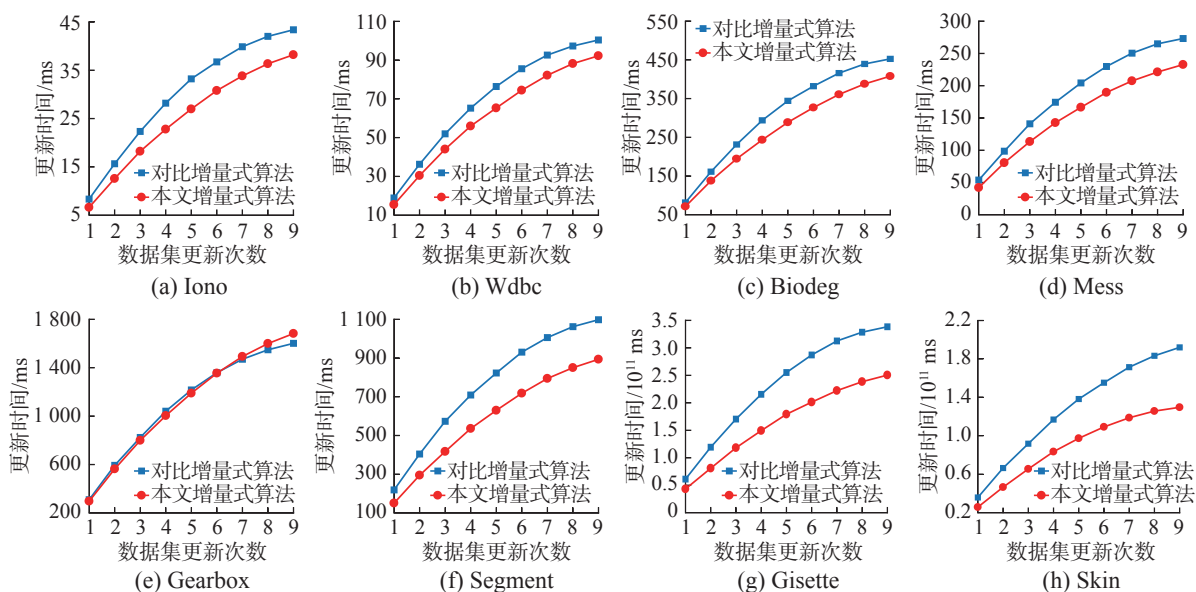


图 6 不同数据量减少时邻域多粒度粗糙集更新用时比较

Fig. 6 Comparison of updating time of neighborhood multi-granulation rough sets with different data reduction

图5给出了信息系统论域增加不同数据量时2种算法的更新时间比较,图6给出了信息系统

论域减少不同数据量时2种算法的更新时间比较。观察图5可以发现,随着数据集更新次数的

增加,即每次增加的数据量逐渐增大,本文增量式算法在所有数据集下的更新时间均小于文献[26]的对比增量式算法,并且效率的差距逐渐拉大。观察图6可以发现,随着数据集更新次数的增加,即每次减少的数据量逐渐增大,除数据集Gearbox的更新时间相当外,本文增量式算法在其余数据集的更新时间均小于文献[26]的对比增量式算法,因此综合起来可以看出,对于信息系统不同数据量的变化情形,本文所提出的增量式更新算法具有较高的更新计算效率。

文献[26]所示的邻域多粒度粗糙集增量式更新算法同样是采用矩阵的形式去构造模型的增量式更新,它主要通过矩阵去直接计算邻域多粒度粗糙集的近似集,即通过邻域关系矩阵去表示每个粒度下的邻域关系,然后通过邻域关系矩阵与近似目标集进行矩阵的标准乘运算,来得到最终近似结果的矩阵形式,当信息系统论域发生变化时,通过增量式的方法去更新邻域关系矩阵,然后通过矩阵运算得到最终的更新近似集结果,在这一系列的增量式计算过程中,其中涉及到了较多的矩阵运算,因此其计算复杂度会比较高。而本文提出的增量式更新方法,虽然也同样采用了矩阵的形式,但是本文没有利用矩阵去表示邻域关系,而是去表示邻域粒与近似集之间的关系,屏蔽了邻域类与近似集之间的运算过程,该过程运算量是最大的,对于信息系统变化后,对近似关系矩阵进行更新,可以减少很多计算量,提高了计算效率,因此本文算法的增量式更新效率会更高一些。此外,本文增量式算法与对比增量式算法的实验结果中,出现了一小部分两种算法计算效率相近的情形,可能是由于当时数据集的更新,使得对于本文算法中近似关系矩阵的更新量较大,导致了耗时较长,从而与对比算法的用时比较接近。

4.4 实验总结

综合所有的实验结果和实验分析可以看出,相比较于邻域多粒度粗糙集非增量式更新算法,本文所提出的增量式更新算法具有非常高的更新效率,与文献[26]提出的增量式更新算法相比,本文的增量式更新算法其性能具有明显的提升,因此证明了本文邻域多粒度粗糙集增量式更新算法具有更高的优越性。

5 结束语

邻域多粒度粗糙集模型是当前粗糙集理论的重要研究分支,本文针对邻域多粒度粗糙集模型

对象变化的情形提出了一种增量式更新算法。文中首先利用矩阵对邻域多粒度粗糙集模型中的近似关系进行表示,定义了近似关系矩阵的概念,理论分析表明,近似关系矩阵是对整个邻域多粒度粗糙集中邻域类与近似集之间细节关系的直观展示,然后针对信息系统对象增加和对象减少的情形,提出了近似关系矩阵的增量式更新,最后利用近似关系矩阵的增量式更新提出了邻域多粒度粗糙集的增量式更新算法,实验验证了本文算法的有效性和优越性。根据本文提出的邻域多粒度粗糙集增量式更新算法,接下来可以针对增量式属性约简进行进一步地研究。

参考文献:

- [1] PAWLAK Z. Rough sets[J]. International journal of computer & information sciences, 1982, 11(5): 341–356.
- [2] BARMAN B, PATRA S. Variable precision rough set based unsupervised band selection technique for hyperspectral image classification[J]. Knowledge-based systems, 2020, 193: 105414.
- [3] 盛魁, 王伟, 卞显福, 等. 混合数据的邻域区分度增量式属性约简算法[J]. 电子学报, 2020, 48(4): 682–696.
SHENG Kui, WANG Wei, BIAN Xianfu, et al. Neighborhood discernibility degree incremental attribute reduction algorithm for mixed data[J]. Acta electronica sinica, 2020, 48(4): 682–696.
- [4] QIAN Yuhua, LIANG Jiye, YAO Yiyu, et al. MGRS: a multi-granulation rough set[J]. Information sciences, 2010, 180(6): 949–970.
- [5] QIAN Yuhua, ZHANG Hu, SANG Yanli, et al. Multi-granulation decision-theoretic rough sets[J]. International journal of approximate reasoning, 2014, 55(1): 225–237.
- [6] QIAN Jin, LIU Caihui, YUE Xiaodong. Multigranulation sequential three-way decisions based on multiple thresholds[J]. International journal of approximate reasoning, 2019, 105: 396–416.
- [7] XU Yi. Multigranulation rough set model based on granulation of attributes and granulation of attribute values[J]. Information sciences, 2019, 484: 1–13.
- [8] LIN Guoping, QIAN Yuhua, LI Jinjin. NMGRS: Neighborhood-based multigranulation rough sets[J]. International journal of approximate reasoning, 2012, 53(7): 1080–1093.
- [9] 骆公志, 钱佳丽. 基于多阈值的变精度邻域多粒度粗糙决策方法[J]. 计算机应用研究, 2018, 35(10): 2943–2946, 3010.
LUO Gongzhi, QIAN Jiali. Neighborhood multi-granulation rough set based on multi-threshold for variable precision

- sion decisions[J]. Application research of computers, 2018, 35(10): 2943–2946,3010.
- [10] 刘丹, 徐立新, 李敬伟. 不完备邻域多粒度决策理论粗糙集与三支决策[J]. 计算机应用与软件, 2019, 36(5): 145–157.
- LIU Dan, XU Lixin, LI Jingwei. Incomplete neighborhood multi-granulation decision-theoretic rough set and three-way decision[J]. Computer applications and software, 2019, 36(5): 145–157.
- [11] 程映, 刘勇. 基于邻域多粒度粗糙集的知识发现模型[J]. 计算机科学, 2019, 46(6): 224–230.
- CHENG Yi, LIU Yong. Knowledge discovery model based on neighborhood multi-granularity rough sets[J]. Computer science, 2019, 46(6): 224–230.
- [12] LUO Chuan, LI Tianrui, YI Zhang, et al. Matrix approach to decision-theoretic rough sets for evolving data[J]. Knowledge-based systems, 2016, 99: 123–134.
- [13] WEI Wei, WU Xiaoying, LIANG Jiye, et al. Discernibility matrix based incremental attribute reduction for dynamic data[J]. Knowledge-based systems, 2018, 140: 142–157.
- [14] SHU Wenhao, QIAN Wenbin, XIE Yonghong. Incremental feature selection for dynamic hybrid data using neighborhood rough set[J]. Knowledge-based systems, 2020, 194: 105516.
- [15] HUANG Yanyong, GUO Kejun, LI Zhong, et al. Matrix representation of the conditional entropy for incremental feature selection on multi-source data[J]. Information sciences, 2022, 591: 263–286.
- [16] 沈玉峰. 基于矩阵方法的区分度增量式属性约简算法[J]. 计算机应用与软件, 2020, 37(9): 235–245,257.
- SHEN Yufeng. A discrimination degree incremental attribute reduction based on matrix method[J]. Computer applications and software, 2020, 37(9): 235–245,257.
- [17] YANG Xin, YANG Yuxuan, LUO Junfang, et al. A unified incremental updating framework of attribute reduction for two-dimensionally time-evolving data[J]. Information sciences, 2022, 601: 287–305.
- [18] HUANG Qianqian, LI Tianrui, HUANG Yanyong, et al. Incremental three-way neighborhood approach for dynamic incomplete hybrid data[J]. Information sciences, 2020, 541: 98–122.
- [19] 杨臻, 邱保志. 混合信息系统的动态变精度粗糙集模型[J]. 控制与决策, 2020, 35(2): 297–308.
- YANG Zhen, QIU Baozhi. Dynamic variable precision rough set model of mixed information system[J]. Control and decision, 2020, 35(2): 297–308.
- [20] 刘丹, 李敬伟. 基于矩阵的双论域模糊概率粗糙集增量更新算法[J]. 控制与决策, 2021, 36(3): 553–564.
- LIU Dan, LI Jingwei. Incremental updating of fuzzy probability rough sets over two universes based on matrix method[J]. Control and decision, 2021, 36(3): 553–564.
- [21] YANG Xibei, QI Yong, YU Hualong, et al. Updating multigranulation rough approximations with increasing of granular structures[J]. Knowledge-based systems, 2014, 64: 59–69.
- [22] HU Chengxiang, LIU Shixi, LIU Guoxiu. Matrix-based approaches for dynamic updating approximations in multigranulation rough sets[J]. Knowledge-based systems, 2017, 122: 51–63.
- [23] HU Chengxiang, ZHANG Li. Efficient approaches for maintaining dominance-based multigranulation approximations with incremental granular structures[J]. International journal of approximate reasoning, 2020, 126: 202–227.
- [24] HU Chengxiang, ZHANG Li, WANG Bangjun, et al. Incremental updating knowledge in neighborhood multi-granulation rough sets under dynamic granular structures[J]. Knowledge-based systems, 2019, 163: 811–829.
- [25] 徐怡, 孙伟康, 王泉. 邻域多粒度粗糙集知识更新增量算法[J]. 小型微型计算机系统, 2020, 41(5): 908–918.
- XU Yi, SUN Weikang, WANG Quan. Incremental algorithm for knowledge updating of neighborhood multi-granulation rough set[J]. Journal of Chinese computer systems, 2020, 41(5): 908–918.
- [26] HU Chengxiang, ZHANG Li. A dynamic framework for updating neighborhood multigranulation approximations with the variation of objects[J]. Information sciences, 2020, 519: 382–406.

作者简介:



陈宝国, 副教授, 安徽省人工智能学会理事、计算机协会会员, 主要研究方向为粗糙集、粒计算、数据挖掘。主持省级以上科研项目 2 项, 参与国家自然科学基金项目 1 项。近年以第一作者发表学术论文 10 余篇。



邓明, 教授, 博士, 安徽省计算机学会理事、安徽省人工智能协会会员, 主要研究方向为矿山安全信息处理、智能数据处理。主持省级以上科研项目 4 项, 参与国家自然科学基金、国家科技支撑项目 2 项, 授权发明专利 2 项, 获省级科技成果 2 项, 获安徽省第七届自然科学优秀学术论文二等奖 1 篇。近年以第一作者发表学术论文 20 余篇, 出版学术专著 1 部。